

TF-MIDAS: a transfer function based mixed-frequency model

Nicolás Bonino-Gayoso* and Alfredo Garcia-Hiernaux†

Version: June 29, 2020

Abstract

This paper tackles the mixed-frequency modeling problem from a new perspective. Instead of drawing upon the common distributed lag polynomial model, we use a transfer function representation to develop a new type of models, named TF-MIDAS. We derive the theoretical TF-MIDAS implied by the high-frequency VARMA family models for two common aggregation schemes, flow and stock. This exact correspondence leads to potential gains in terms of nowcasting and forecasting performance against the current alternatives. The estimation of the model proposed is also addressed via its state space equivalent form. A Monte Carlo simulation exercise confirms that TF-MIDAS beats U-MIDAS models (its natural competitor) in terms of out-of-sample nowcasting performance for several data generating high-frequency processes.

Keywords: mixed-frequency models, U-MIDAS, Nowcasting, Forecasting, TF-MIDAS

JEL: C18, C51, C53

1 Introduction

Economic policy-makers, entrepreneurs and investors, among other agents, need to have access to real-time assessments of the state of the economy, along with nowcasts and forecasts of its expected evolution. The sooner they have access to information about the real economic situation, the better prepared they will be to make decisions to update their initial plans. This becomes even more important in non-stable economic and challenging environments like the one we are now facing with the Covid-19.

Unfortunately, data offered by the System of National Accounts (SNA) is delivered with considerable delay. In the case of European countries, Eurostat provides the value of

*PhD candidate at Universidad Complutense de Madrid; Decon, Universidad de la República, Uruguay. E-mail: nicolas.bonino.gayoso@gmail.com

†DANAE and ICAE, Universidad Complutense de Madrid. E-mail: agarciah@ucm.es

EU and euro area GDP, 70 days after the end of quarter, preceded by a first preliminary estimate and a second estimate 30 and 45 days after the end of quarter, respectively.¹ This delay, as Castle and Hendry (2013) point out, is the consequence of the existence of several difficulties for producing timely and accurate low-frequency aggregates: (i) not all disaggregated data are available when is needed to compute a relevant aggregate; (ii) many disaggregated time series are only preliminary estimates, subject to substantial revisions, so they are not accurate descriptions of the current conditions.

On the other hand, there is a considerable number of short-term economic indicators available at a much earlier stage that could be used to extract information about the state of the economy. Thus we have access to monthly data from consumer surveys or the industrial production index, daily data from financial markets and even more frequently observed variables, such as Google or Twitter trends and mobile phone data. Although not as complete as SNA data, these indicators could anticipate relevant information.

Forecasting models usually require to use data observed at the same frequency, which represents a setback when using a more complete information set provided by mixed-frequency variables. Several solutions have been developed to try to overcome this problem. In the following lines we will present the most common ones.

Throughout the paper we follow the notation used in Foroni et al. (2015). A high-frequency (HF) indicator is denoted by letter x . Let t be the time index for this variable x , $t = 1, \dots, T$ (e.g., although not exclusively, months), being T the last period for which data of variable x is available. The lag operator for this HF indicator is denoted by L . If x_t is the monthly industrial production index used to nowcast quarterly GDP, then Lx_t will be the value of this index corresponding to the previous month.

Let y be the low-frequency (LF) variable that is aimed to be nowcast, sampled at periods denoted by time index $t_q = 1, \dots, T_q$ (e.g., although not exclusively, quarters), being T_q the last period for which data of variable y is available. Usually, $T \geq k * T_q$, as observations of HF indicators are available earlier than LF ones. Past realizations of the LF variable will be denoted by the lag operator Z , where $Z = L^k$. So, if y_{t_q} is quarterly GDP, then Zy_{t_q} will be GDP of the previous quarter.

The HF indicator x is sampled k times between samples of y . For example, for quarterly GDP and a monthly indicator, $k = 3$.

Finally, aggregated values of the HF indicator are denoted by x_{t_q} , e.g., quarterly aggregation of the monthly industrial production index. Both, the target variable y and the indicator x are assumed to be stationary, so these variables often correspond to a (log) differenced version of some raw series z .

The way to extract information from the available indicators is not straightforward and there are several methodologies, with different levels of complexity, to address this task. Several classes of models have been proposed to work explicitly with mixed-frequency datasets, most widely used being MIDAS (MIXed Data Sampling) family of models.

¹In the case of USA, the Bureau of Economic Analysis (BEA) releases a first estimate of GDP 28 days after the end of quarter, followed by a second and a third estimate 60 and 90 days after the end of quarter, respectively. In Latin American countries the schedule is similar. In Uruguay, for example, GDP value is released by the Central Bank 75 days after the end of quarter.

MIDAS models (Ghysels et al., 2002, 2003, 2006) are defined in terms of a Distributed Lag (DL) polynomial, explicitly modelling the relationship between variables observed at different frequencies. In order to keep parsimony, standard MIDAS models are defined in terms of a short number of parameters. This kind of model has been applied to nowcast GDP, private consumption and corporate bond spreads, among other variables (Ghysels et al., 2007; Clements and Galvão, 2008, 2009; Bai et al., 2013; Schumacher, 2014; Duarte et al., 2017).

A specific variation of the standard model, known as Unrestricted MIDAS (U-MIDAS), is introduced by Foroni et al. (2015). Based on a series of simulation exercises, these authors state that U-MIDAS' nowcasting accuracy outperforms that one of standard MIDAS, when the difference in sampling frequencies is not large, specially for monthly to quarterly frequency, as it is usually the case of macroeconomic nowcasting.

In this paper, we apply a transfer function representation to the DL polynomial, deriving a new type of mixed-frequency models, named Transfer Function-MIDAS or, simply, TF-MIDAS. We consider that the representation of the classical MIDAS model in terms of ratios of finite lag polynomials is a more appropriate way to model mixed-frequency data, as it is a more general alternative than current DL models. Moreover, a formal methodology of specification and estimation was already developed and tested long time ago (Box and Jenkins, 1976).

The contribution of the paper is three-fold. First we present a new type of models directly derived from a potential multivariate HF relationship between the aggregate and the indicators and the two most common aggregation schemes (skip-sampling and addition). Second, the estimation of these models is addressed through the formulation to its equivalent state space form. Third, Monte Carlo simulation exercises reveal that TF-MIDAS beats commonly employed U-MIDAS models in terms of out-of-sample nowcasting performance. Working with simulated data allows us to consider different variants of Data Generating Processes (DGP) and so, identify for the specific DGPs for which the advantage of TF-MIDAS is significantly greater.

We conclude that TF-MIDAS presents an overall better nowcasting performance than U-MIDAS. In the case of a HF-VAR (1) process the out-of-sample nowcasting performance of both models is practically the same, whereas in the case of a VMA(1) and VMA(3) processes TF-MIDAS outstands as a more accurate nowcasting model. These results are robust to different parameter specifications.

Given that the paper is already dense, long and discusses many aspects of the model, for the sake of space it does not deal with the impact of misspecification in the nowcast/forecast or the study of a real case. In fact, these two interesting aspects are developed in a different paper with a detailed real data analysis that brings the opportunity to deepen into those results.

The paper is organized in five sections, including the present introduction. In section 2, MIDAS models are briefly reviewed, specially focusing on one specific variant of these models known as U-MIDAS. Section 3 is dedicated to introduce our proposed model, TF-MIDAS. Theoretical elements and the estimation are discussed in this section. In section 4, the simulation design and the models under comparison are detailed and results of relative out-of-sample nowcasting performance of TF-MIDAS model are examined. We

determine for which kind of processes TF-MIDAS shows a significantly better relative performance and so it would be advisable to apply it. Finally, Section 5 summarizes the main results and concludes.

2 Review of MIDAS models

Prior to review the specific U-MIDAS model, it is worth presenting a brief summary of the original MIDAS (MIXed DATA Sampling) model, which was introduced for the first time in two working papers (Ghysels et al., 2002, 2003) and later in a published paper (Ghysels et al., 2006).

In MIDAS models the response of the LF variable to a HF explicative variable is modeled through a DL polynomial and special attention is paid to parsimony. In order to avoid the so-called “parameter proliferation” problem, lag coefficients are not free, but are defined as a function of a vector of few parameters, θ , known as hyperparameters.

Ghysels et al. (2002) propose the following DL specification for a MIDAS model with N explicative variables:

$$\omega(L)y_t = \beta_0 + \sum_{i=1}^N \beta_i B_i(L; \theta) x_{i,t-1} + \epsilon_t \quad t = k, 2k, \dots, Tk, \quad (1)$$

where $B_i(L; \theta) = \sum_{j=0}^K b_i(j; \theta) L^j$ and $\omega(L) = \omega_0 + \omega_1 L + \dots + \omega_{k-1} L^{k-1}$ determines the aggregation scheme.²

Function $b(j, \theta)$, a component of the lag polynomial, is used to model the weights assigned to each lag of the HF indicator. This function depends on the indicator’s period, j , and the hyperparameters. An overview of different weighting functions proposed so far in the literature is provided in Ghysels (2014), the most popular being Exponential Almon and Beta.³

There are several variations built upon the basic MIDAS model. A detailed summary of the main features of these variations can be found in Foroni and Marcellino (2013).

²In the case of a stock variable, it would be $\omega_0 = 1, \omega_1 = \dots = \omega_{k-1} = 0$. In the case of a flow variable, it would be $\omega_0 = \omega_1 = \dots = \omega_{k-1} = 1$, if the values are aggregated by addition, or $\omega_0 = \omega_1 = \dots = \omega_{k-1} = 1/k$, if the values are aggregated by average.

³The Exponential Almon weighting function was proposed in Ghysels et al. (2005) and it has the following expression, with Q shape parameters:

$$b(j; \theta) = \frac{\exp(\theta_1 j + \dots + \theta_Q j^Q)}{\sum_{j=0}^K \exp(\theta_1 j + \dots + \theta_Q j^Q)} \quad (2)$$

Beta weighting function, proposed for the first time in Ghysels et al. (2003), includes only two shape parameters:

$$b(j; \theta_1, \theta_2) = \frac{f(\frac{j}{K}; \theta_1, \theta_2)}{\sum_{j=1}^K f(\frac{j}{K}; \theta_1, \theta_2)} \quad (3)$$

where f is the Beta probability density function.

2.1 U-MIDAS model

Foroni et al. (2012, 2015) propose a variant of MIDAS model known as U-MIDAS, which does not employ functional distributed lag polynomials to model the relationship between x and y .

When the difference in sampling frequencies is not large, the risk of falling into the curse of dimensionality becomes less relevant and so it does the need to resort to functional DL polynomials.

The U-MIDAS model based on a linear lag polynomial is defined by:

$$C(L^k)\omega(L)y_t = \delta_1(L)x_{1,t-1} + \dots + \delta_N(L)x_{N,t-1} + \epsilon_t \quad t = k, 2k, \dots, Tk, \quad (4)$$

where $C(L^k) = 1 - c_1L^k - \dots - c_cL^{kc}$; $\delta_j(L) = \delta_{j,0} + \delta_{j,1}L + \dots + \delta_{j,v_j}L^{v_j}$, $j = 1, \dots, N$; x_j , $j = 1, \dots, N$ are the explanatory variables, each of them affecting y up to lag v_j . Foroni et al. (2015) state that notwithstanding the error term ϵ_t has in general a moving average structure, i.e., $\epsilon_t = C(L^k)a_t$, where $a_t \sim iid(0, \sigma^2)$, they suggest to consider an AR approximation for simplicity. Finally, $\omega(L) = \omega_0 + \omega_1L + \dots + \omega_{k-1}L^{k-1}$ determines the aggregation scheme.

Foroni et al. (2015) state that basic MIDAS model can be thought of as nested in U-MIDAS specification, because it is the result of imposing a particular dynamic pattern on it. An important computational advantage of U-MIDAS model over the basic MIDAS model is that it can be estimated by simple OLS, as long as lag orders c and v_j are long enough to make the error term ϵ_t uncorrelated.

Once the model is fitted, nowcasting for y_{Tk+k} conditional on information available at period $Tk + k - 1$ is expressed as:

$$\hat{y}_{Tk+k|Tk+k-1} = (\hat{c}_1L^k + \dots + \hat{c}_cL^{kc})y_{Tk+k} + \hat{\delta}_1(L)x_{1,Tk+k-1} + \dots + \hat{\delta}_N(L)x_{N,Tk+k-1}. \quad (5)$$

3 TF-MIDAS

In the literature about mixed-frequency models, MIDAS models are represented in terms of a DL polynomial expression, as they were shown in the previous section. In this section we present an alternative representation based on a Transfer Function (TF) model, which we refer to as TF-MIDAS. We also derive the new proposed model from a general linear dynamic model and compare it, from a theoretically viewpoint, with a U-MIDAS model.

3.1 Transfer Function Mixed Data Sampling model (TF-MIDAS)

We build on the hypothesis that TF-MIDAS is a more appropriate way to model mixed-frequency data as it is a more general representation than previous DL models in two ways: (i) removing the truncation error, and (ii) including an MA component to the model. We will pay special attention to this two new features in Subsection 3.3.

We define the general TF-MIDAS model with N indicators using two equations. First, the equation that models the relation between y and x :

$$\omega(L)y_t = \beta_0 + \sum_{i=1}^N \sum_{j=1}^k \frac{a_{i,j}(Z)}{b_{i,j}(Z)} x_{i,t-j} + \epsilon_t. \quad (6a)$$

Second, the equation for the noise:

$$\phi(Z)\epsilon_t = \theta(Z)a_t, \quad (6b)$$

where $a_j(Z)$ and $b_j(Z)$ are finite lag polynomials, $Z = L^k$; $\phi(Z)$ and $\theta(Z)$ are polynomials of order p and q , respectively, with the usual properties for stationary and invertible ARMA processes (see, e.g., Box and Jenkins, 1976).

To estimate TF-MIDAS models, we rearrange the dataset in the following way:

$$\begin{array}{c|cccc|cccc|c|cccc} y_k^* & x_{1,k-1} & x_{1,k-2} & \dots & x_{1,0} & x_{2,k-1} & x_{2,k-2} & \dots & x_{2,0} & \dots & x_{N,k-1} & x_{N,k-2} & \dots & x_{N,0} \\ y_{2k}^* & x_{1,2k-1} & x_{1,2k-2} & \dots & x_{1,k} & x_{2,2k-1} & x_{2,2k-2} & \dots & x_{2,k} & \dots & x_{N,2k-1} & x_{N,2k-2} & \dots & x_{N,k} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ y_{Tk}^* & x_{1,Tk-1} & x_{1,Tk-2} & \dots & x_{1,Tk-k} & x_{2,Tk-1} & x_{2,Tk-2} & \dots & x_{2,Tk-k} & \dots & x_{N,Tk-1} & x_{N,Tk-2} & \dots & x_{N,Tk-k} \end{array}$$

where $y_i^* = \omega(L)y_t$ refers to the LF aggregated value of the endogenous variable y ; $x_{1,t}$, $x_{2,t}$, ..., $x_{N,t}$ are the HF values of the exogenous variables x_1 , x_2 , ..., x_N .

Throughout the remaining paper, and for the lack of simplicity, in the theoretical expressions we use only one indicator (x) in the model. This does not affect in any way the results and conclusions derived from the analysis, beyond the computational issue that will be discussed later on.

For monthly-quarterly data (i.e., $k = 3$) and only one indicator the previous equations becomes:

$$\omega(L)y_t = \beta_0 + \frac{a_1(Z)}{b_1(Z)} x_{t-1} + \frac{a_2(Z)}{b_2(Z)} x_{t-2} + \frac{a_3(Z)}{b_3(Z)} x_{t-3} + \epsilon_t \quad (7a)$$

$$\phi(Z)\epsilon_t = \theta(Z)a_t, \quad (7b)$$

where x_{t-1} is a vector made up of the second monthly observation of the current quarter, x_{t-2} is made up of the first monthly observation of the current quarter, and x_{t-3} is made up of the third monthly observation of the previous quarter. In this context, the data is organized in the following way:

$$\begin{array}{c|cccc} y_3^* & x_{1,2} & x_{1,1} & x_{1,0} & \\ y_6^* & x_{1,5} & x_{1,4} & x_{1,3} & \\ \dots & \dots & \dots & \dots & \\ y_{3T}^* & x_{1,3T-1} & x_{1,3T-2} & x_{1,3T-3} & \end{array}$$

The vector of aggregated values of the endogenous variable y is named y_{t_q} ; the first, second and third vector of the exogenous variable are named x_{1,t_q} , x_{2,t_q} and x_{3,t_q-1} , respectively.

3.2 Rationale behind TF-MIDAS models

In this section we derive the exact TF-MIDAS models corresponding to different potential VARMA processes, which represent the relationship between the aggregate and the indicator in the HF. We do so for the usual aggregation schemes skip-sampling and addition/mean, corresponding to stock and flow aggregates, respectively.

We assume that HF observations of variables y and x are generated by a VARMA(p, q) process:

$$\begin{bmatrix} \Phi_{1,1} & \Delta_{1,2} \\ 0 & \Phi_{2,2} \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} \Theta_{1,1} & \Pi_{1,2} \\ 0 & \Theta_{2,2} \end{bmatrix} \begin{bmatrix} e_{y,t} \\ e_{x,t} \end{bmatrix}, \quad (8)$$

where $\Phi_{1,1} = 1 + \phi_{l,1}L + \phi_{l,2}L^2 + \dots + \phi_{l,p}L^p$; $\Phi_{2,2} = 1 + \phi_{h,1}L + \phi_{h,2}L^2 + \dots + \phi_{h,p}L^p$; $\Delta_{1,2} = \delta_{l,1}L + \delta_{l,2}L^2 + \dots + \delta_{l,p}L^p$; $\Theta_{1,1} = 1 + \theta_{l,1}L + \theta_{l,2}L^2 + \dots + \theta_{l,q}L^q$; $\Theta_{2,2} = 1 + \theta_{h,1}L + \theta_{h,2}L^2 + \dots + \theta_{h,q}L^q$; $\Pi_{1,2} = \pi_{l,1}L + \pi_{l,2}L^2 + \dots + \pi_{l,q}L^q$. Throughout the whole paper we assume that $\Phi_{i,i}$ and $\Theta_{i,i}$ polynomials hold the usual conditions for stationarity and invertibility. We employ the family models represented in (8) because it is commonly used in the literature (see, e.g., Ghysels and Valkanov, 2006; Foroni et al., 2015).

We illustrate TF-MIDAS models with three HF DGPs: (i) VAR(1); (ii) VMA(1); (iii) VMA(3), and two aggregation methods: 1) skip-sampling, and 2) addition.⁴ The next subsection derives an exact TF-MIDAS representation for a VAR(1) DGP, being y a flow variable (i.e., applying the addition as aggregation method). Derivations for the rest of the models are in the Appendix.

3.2.1 TF-MIDAS representation of a HF-VAR(1) and a flow variable

Here we assume monthly values of variables y and x are generated from a VAR(1) process:

$$y_t = \phi_l y_{t-1} + \delta_l x_{t-1} + e_{y,t} \quad (9a)$$

$$x_t = \phi_h x_{t-1} + e_{x,t}. \quad (9b)$$

Shifting (9a), we have:

$$y_{t-k} = \phi_l y_{t-k-1} + \delta_l x_{t-k-1} + e_{y,t-k} \quad t = 1, 2, 3, \dots \quad (10)$$

Considering the previous expression for $k = 1$ and $k = 2$, and substituting them in (9a), yields:

$$y_t = \phi_l^3 y_{t-3} + \delta_l x_{t-1} + \phi_l \delta_l x_{t-2} + \phi_l^2 \delta_l x_{t-3} + \phi_l^2 e_{y,t-2} + \phi_l e_{y,t-1} + e_{y,t} \quad (11)$$

Rearranging terms, we can represent (11) in TF form, expressing the value of y in terms of HF values of variable x and an error term:

$$y_t = \frac{\delta_l}{1 - \phi_l^3 L^3} x_{t-1} + \frac{\phi_l \delta_l}{1 - \phi_l^3 L^3} x_{t-2} + \frac{\phi_l^2 \delta_l}{1 - \phi_l^3 L^3} x_{t-3} + \frac{1}{1 - \phi_l L} e_{y,t} \quad (12)$$

⁴We also develop the theoretical TF-MIDAS representations corresponding to additional VAR(p), VMA(q) and VARMA(p, q) HF processes. The developments are available from the authors upon request.

where $t = 4, 5, 6, \dots$ and in order to obtain the last expression we use the factorization $1 - \phi_l^3 L^3 = (1 - \phi_l L)(1 + \phi_l L + \phi_l^2 L^2)$.

From the expression for y_t in (12) we can deduce equivalent expressions for y_{t-1} and y_{t-2} . Adding these expressions for y_t , y_{t-1} and y_{t-2} , and rearranging terms, we get the aggregated quarterly value of y in the case of a flow variable, denoted by y_t^A :

$$y_t^A = \frac{\delta_l + [\phi_l \delta_l (1 + \phi_l)] L^3}{1 - \phi_l^3 L^3} x_{t-1} + \frac{\delta_l (1 + \phi_l) + (\phi_l^2 \delta_l) L^3}{1 - \phi_l^3 L^3} x_{t-2} + \frac{\delta_l (1 + \phi_l + \phi_l^2)}{1 - \phi_l^3 L^3} x_{t-3} \\ + \frac{1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4}{1 - \phi_l^3 L^3} e_{y,t} \quad t = 4, 5, 6, \dots \quad (13)$$

Finally, from (13) we derive the equation that expresses the LF aggregated value of a flow variable y in terms of the HF values of x :

$$y_{t_q}^A = \frac{\delta_l + [\phi_l \delta_l (1 + \phi_l)] Z}{1 - \phi_l^3 Z} x_{2,t_q} + \frac{\delta_l (1 + \phi_l) + (\phi_l^2 \delta_l) Z}{1 - \phi_l^3 Z} x_{1,t_q} \\ + \frac{\delta_l (1 + \phi_l + \phi_l^2)}{1 - \phi_l^3 Z} x_{3,t_q-1} + \frac{1}{1 - \phi_l^3 Z} \eta_{t_q} \quad t_q = 2, 3, 4, \dots \quad (14)$$

where $y_{t_q}^A$ is the aggregated quarterly value of y , which is a flow variable, i.e. $y_{t_q}^A = y_t + y_{t-1} + y_{t-2}$ for $t = 3 * t_q$; x_{2,t_q} is the second monthly value of x for current quarter; x_{1,t_q} is the first monthly value of x for current quarter; x_{3,t_q-1} is the last monthly value of x for the previous quarter; and $\eta_{t_q} = [1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4] e_{y,t}$.

By construction, η_{t_q} is an autocorrelated noise. We now calculate its principal moments as a function of $e_{y,t}$. Expected value:

$$E[(1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4) e_{y,t}] = 0, \quad (15)$$

variance:

$$V[(1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4) e_{y,t}] \\ = (3 + 4\phi_l + 5\phi_l^2 + 4\phi_l^3 + 3\phi_l^4) V[e_{y,t}], \quad (16)$$

and, covariances:

$$Cov[(1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4) e_{y,t}, \\ (1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4) e_{y,t-3}] \\ = \phi_l (1 + \phi_l)^2 V[e_{y,t}] \quad (17a)$$

$$Cov[(1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4) e_{y,t}, \\ (1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4) e_{y,t-3k}] \\ = 0, \quad k = 2, 3, 4, \dots \quad (17b)$$

From (17a-17b) the autocorrelations are:

$$\rho_1 = \frac{\phi_l(1 + \phi_l)^2}{3 + 4\phi_l + 5\phi_l^2 + 4\phi_l^3 + 3\phi_l^4} \quad (18a)$$

$$\rho_k = 0, \quad k = 2, 3, 4, \dots \quad (18b)$$

Therefore, η_{t_q} presents a MA(1) structure, which can be written as:

$$\eta_{t_q} = (1 + \psi Z) e_{y^A, t_q}, \quad (19)$$

where e_{y^A, t_q} is a white noise process.⁵

Similarly derived in the Appendix, Table 1 shows a TF-MIDAS representation for each one of the three DGPs here considered and both, a stock and a flow y variable.

3.3 Theoretical comparison between TF-MIDAS and U-MIDAS models

Previous transfer function (7a-7b) can also be regarded as a DL model under two restrictions: (i) $\theta(Z) = 1$ and (ii) $b_i(Z) = \phi(Z)$, where lag operator Z has its roots outside the unit circle, in order to ensure stationarity.

Substituting equation (7b) into (7a) and multiplying both sides of the resulting equation by $\phi(Z)$, yields:

$$\phi(Z)\omega(L)y_t = \beta_0^* + \omega_1(Z) x_{t-1} + \omega_2(Z) x_{t-2} + \omega_3(Z) x_{t-3} + a_t \quad (20)$$

where $\beta_0^* = \phi(Z)\beta_0$. The expression obtained is equivalent to an U-MIDAS model. It implies so that lags of the LF variable are included in the model.

We can reaffirm this equivalence by considering Equation (8) in Foroni et al. (2015), which states:

$$c(L^k)\omega(L)y_t = \delta_1(L)x_{1,t-1} + \dots + \delta_N(L)x_{N,t-1} + \epsilon_t \quad t = k, 2k, 3k, \dots \quad (21)$$

When dividing both sides of the equation by $c(L^k)$ we obtain an equivalent TF-MIDAS expression:

$$\omega(L)y_t = \frac{\delta_1(L)}{c(L^k)}x_{1,t-1} + \dots + \frac{\delta_N(L)}{c(L^k)}x_{N,t-1} + \frac{1}{c(L^k)}\epsilon_t \quad t = k, 2k, 3k, \dots \quad (22)$$

This equivalence between TF-MIDAS and U-MIDAS stands only if the two restrictions mentioned earlier are satisfied. Otherwise, we could derive from equations (7a-7b) this another equivalent expression for TF-MIDAS model:

$$\omega(L)y_t = \beta_0 + \frac{a_1(Z)}{b_1(Z)} x_{t-1} + \frac{a_2(Z)}{b_2(Z)} x_{t-2} + \frac{a_3(Z)}{b_3(Z)} x_{t-3} + \frac{\theta(Z)}{\phi(Z)} a_t \quad (23)$$

⁵The value of ψ in (19) can be calculated straightforwardly by using the ACF for a MA(1) model: $\rho_1 = \psi/(1 + \psi^2)$ and $\rho_k = 0$ for $k = 2, 3, \dots$. Solving the ρ_1 equation for ψ will return two solutions, from where we consider the one satisfying the invertibility condition: $|\psi| < 1$.

Table 1: **Theoretical TF-MIDAS models corresponding to selected HF-DGP and different aggregation schemes**

HF-DGP VAR(1)	
	$y_t = \phi_l y_{t-1} + \delta_l x_{t-1} + e_{y,t}$ $x_t = \phi_h x_{t-1} + e_{x,t}$
TF-MIDAS Stock	$y_{t_q}^A = \frac{\delta_l}{1 - \phi_l^3 Z} x_{2,t_q} + \frac{\phi_l \delta_l}{1 - \phi_l^3 Z} x_{1,t_q} + \frac{\phi_l^2 \delta_l}{1 - \phi_l^3 Z} x_{3,t_q-1} + \zeta_{t_q}$ $(1 - \phi_l^3 Z) \zeta_{t_q} = e_{y^A,t_q}$
TF-MIDAS Flow	$y_{t_q}^A = \frac{\delta_l + [\phi_l \delta_l (1 + \phi_l)] Z}{1 - \phi_l^3 Z} x_{2,t_q} + \frac{\delta_l (1 + \phi_l) + (\phi_l^2 \delta_l) Z}{1 - \phi_l^3 Z} x_{1,t_q} + \frac{\delta_l (1 + \phi_l + \phi_l^2)}{1 - \phi_l^3 Z} x_{3,t_q-1} + \zeta_{t_q}$ $(1 - \phi_l^3 Z) \zeta_{t_q} = (1 + \psi Z) e_{y^A,t_q}$
HF-DGP VMA(1)	
	$y_t = e_{y,t} + \theta_l e_{y,t-1} + \delta_l e_{x,t-1}$ $x_t = e_{x,t} + \theta_h e_{x,t-1}$
TF-MIDAS Stock	$y_{t_q}^A = \frac{\delta_l}{1 + \theta_h^3 Z} x_{2,t_q} - \frac{\delta_l \theta_h}{1 + \theta_h^3 Z} x_{1,t_q} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 Z} x_{3,t_q-1} + \zeta_{t_q}$ $\zeta_{t_q} = e_{y^A,t_q}$
TF-MIDAS Flow	$y_{t_q}^A = \frac{\delta_l + \delta_l \theta_h (\theta_h - 1) Z}{1 + \theta_h^3 Z} x_{2,t_q} + \frac{\delta_l (1 - \theta_h) + \delta_l \theta_h^2 Z}{1 + \theta_h^3 Z} x_{1,t_q} + \frac{\delta_l (1 - \theta_h + \theta_h^2)}{1 + \theta_h^3 Z} x_{3,t_q-1} + \zeta_{t_q}$ $\zeta_{t_q} = (1 + \psi Z) e_{y^A,t_q}$
HF-DGP VMA(3)	
	$y_t = e_{y,t} + \theta_l e_{y,t-3} + \delta_l e_{x,t-1}$ $x_t = e_{x,t} + \theta_h e_{x,t-1}$
TF-MIDAS Stock	$y_{t_q}^A = \frac{\delta_l}{1 + \theta_h^3 Z} x_{2,t_q} - \frac{\delta_l \theta_h}{1 + \theta_h^3 Z} x_{1,t_q} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 Z} x_{3,t_q-1} + \zeta_{t_q}$ $\zeta_{t_q} = (1 + \theta_l Z) e_{y^A,t_q}$
TF-MIDAS Flow	$y_{t_q}^A = \frac{\delta_l + \delta_l \theta_h (\theta_h - 1) Z}{1 + \theta_h^3 Z} x_{2,t_q} + \frac{\delta_l (1 - \theta_h) + \delta_l \theta_h^2 Z}{1 + \theta_h^3 Z} x_{3,t_q} + \frac{\delta_l (1 - \theta_h + \theta_h^2)}{1 + \theta_h^3 Z} x_{1,t_q-1} + \zeta_{t_q}$ $\zeta_{t_q} = (1 + \theta_l Z) e_{y^A,t_q}$

Multiplying both sides of the previous equation by $\phi(Z)$ and rewriting each polynomial quotient as an infinite series yields the following alternative expression for the TF-MIDAS model:

$$\phi(Z)\omega(L)y_t = \beta_0^* + a_1^*(Z)x_{t-1} + a_2^*(Z)x_{t-2} + a_3^*(Z)x_{t-3} + \theta(Z)a_t \quad (24)$$

where $\beta_0^* = \phi(Z)\beta_0$; $a_i^*(Z) = \sum_{j=0}^{\infty} a_{i,j} Z^j$, $i = 1, 2, 3, \dots$.

Therefore using U-MIDAS model when the two aforementioned conditions are not satisfied implies truncating the lag polynomials in the previous expression, and therefore working with an approximation of the true model. How good this approximation results depends on the specific characteristics of the unobservable HF-DGP.

For example, consider the following TF-MIDAS model:

$$\omega(L)y_t = \beta_0 + \frac{a_{0,1} + a_{1,1}Z}{1 - b_1Z}x_{t-1} + \frac{a_{0,2} + a_{1,2}Z}{1 - b_2Z}x_{t-2} + \frac{a_{0,3} + a_{1,3}Z}{1 - b_3Z}x_{t-3} + \epsilon_t \quad (25)$$

Assuming $0 < b_j < 1$, the quotient $(a_{0,j} + a_{1,j}Z)/(1 - b_jZ)$ can be expanded to $a_{0,j} + (a_{1,j} + a_{0,j}b_j)Z + \delta_j(a_{1,j} + a_{0,j}b_j)Z^2 + b_j^2(a_{1,j} + a_{0,j}b_j)Z^3 + \dots$, which is a polynomial in Z of infinite order. So, the model in TF representation can be rewritten as:

$$\omega(L)y_t = \beta_0 + \beta_1x_{t-1} + \beta_2x_{t-2} + \beta_3x_{t-3} + \dots + \epsilon_t, \quad t = 3, 6, \dots, 3T, \quad (26)$$

where $\beta_n = a_{0,n}$, for $n = 1, 2, 3$; $\beta_{n+3} = (a_{1,n} + a_{0,n}b_n)$, for $n = 1, 2, 3$; $\beta_{n+3i+3} = b_n^i(a_{1,n} + a_{0,n}b_n)$, for $n = 1, 2, 3, \dots$ and $i = 1, 2, 3, \dots$

If instead of the TF representation, we opt for the DL one, it will be specified in terms of a finite-order polynomial, e.g., a third-order polynomial:

$$\omega(L)y_t = \beta_0 + \beta_1x_{t-1} + \beta_2x_{t-2} + \beta_3x_{t-3} + \epsilon_t. \quad (27)$$

Besides the difference in the specification of the error term structure, a crucial difference between both models is the truncation of the lag polynomial in the DL representation. This truncation implies discarding the terms of higher order in the lag polynomial. In our example, it implies to leave out the terms of forth and higher order: $\beta_4x_{t-4} + \beta_5x_{t-5} + \dots$

For some HF-DGPs, the use of a DL approximation will result then in different nowcasts and forecasts with respect to the ones obtained with a TF representation.

In turn, although the TF-MIDAS model is more accurate and parsimonious, it has to be estimated by non-linear techniques, such as Maximum-Likelihood (ML). This, of course, represents a greater computational challenge than fitting U-MIDAS by least squares (LS) procedures. We discuss this issue in the next subsection.

3.4 TF-MIDAS estimation

The parameters in TF-MIDAS are estimated by exact ML. To do so, we rewrite TF-MIDAS in its equivalent state space formulation and run the Kalman Filter equations

for the estimation. As an example, we will consider in this section the theoretical TF-MIDAS that corresponds to a HF-DGP VMA(1) model with a flow aggregation scheme. HF and aggregated LF models can be found in previous Table 1. We choose this model and the flow aggregation scheme as both lead to a more complex TF-MIDAS and it better illustrates the estimation problem.

As shown in Table 1 and the Appendix, a HF-DGP VMA(1) model with a flow aggregation scheme yields the following TF-MIDAS:

$$y_{tq}^A = \frac{a_{0,1} + a_{1,1}Z}{1 - b_1Z}x_{2,tq} + \frac{a_{0,2} + a_{1,2}Z}{1 - b_2Z}x_{1,tq} + \frac{a_{0,3} + a_{1,3}Z}{1 - b_3Z}x_{3,tq-1} + \zeta_{tq}, \quad (28a)$$

$$\zeta_{tq} = (1 + \psi Z)e_{y^A,tq}, \quad (28b)$$

where y_{tq}^A is the LF variable and $x_{i,tq}$ represents the HF indicator, arranged as in Section 3.1. This model can be written in its equivalent state space innovations form as:

$$\mathbf{z}_{tq+1} = \mathbf{\Phi}\mathbf{z}_{tq} + \mathbf{\Gamma}\mathbf{x}_{tq} + \mathbf{E}e_{y^A,tq} \quad (29a)$$

$$y_{tq}^A = \mathbf{H}\mathbf{z}_{tq} + \mathbf{D}\mathbf{x}_{tq} + e_{y^A,tq} \quad (29b)$$

where \mathbf{z}_{tq} is a vector of four states that describes the dynamics of the model and $\mathbf{x}_{tq} = [x_{2,tq} \ x_{1,tq} \ x_{3,tq-1}]^\top$ is the input vector that accommodates the information of the indicator. Matrices in the state space form are related to the parameters in (28a-28b) as follows:

$$\mathbf{\Phi} = \begin{bmatrix} \phi_1 & 1 & 0 & 0 \\ -\phi_2 & 0 & 1 & 0 \\ \phi_3 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{E} = \psi \begin{bmatrix} 1 \\ -\phi_1 \\ \phi_2 \\ -\phi_3 \end{bmatrix} \quad \text{and} \quad \mathbf{H} = [1 \ 0 \ 0 \ 0],$$

being $\phi_1 = b_1 + b_2 + b_3$, $\phi_2 = (b_1 + b_2)(b_1 + b_3)(b_2 + b_3)$ and $\phi_3 = b_1b_2b_3$. And, finally,

$$\mathbf{\Gamma} = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 \\ -\Gamma_1(b_2 + b_3) & -\Gamma_2(b_1 + b_3) & -\Gamma_3(b_1 + b_2) \\ \Gamma_1b_2b_3 & \Gamma_2b_1b_3 & \Gamma_3b_1b_2 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{D} = [a_{0,1} \ a_{0,2} \ a_{0,3}],$$

being $\Gamma_1 = a_{1,1} + a_{0,1}b_1$, $\Gamma_2 = a_{1,2} + a_{0,2}b_2$ and $\Gamma_3 = a_{0,3}b_3$, for those related to the input.

It can be seen that solving (29a) recursively and replacing the result by \mathbf{z}_{tq} in (29b) yields the TF-MIDAS in (28a-28b). As ML convergence does sometimes depend on the initial values of the parameters and TF-MIDAS are models with a considerable number of them, we suggest to use a procedure to get consistent estimates for those values prior to the ML estimation. Here we use the procedure by Garcia-Hiernaux et al. (2009). Then, the exact ML is computed using the standard Kalman filter equations for a state space model with stochastic inputs (see, Casals et al., 2016), by iterating on the set of parameters $\{a_{0,1}, a_{1,1}, a_{0,2}, a_{1,2}, a_{0,3}, a_{1,3}, b_1, b_2, b_3, \psi\}$ and keeping the 0s and 1s in the state space matrices representation constrained to its value.⁶

⁶In order to keep focused on the TF-MIDAS discussion, we do not present in this paper the ML function and the Kalman filter equations, as this is standard in the state space models literature. However, for readers who are not familiar with this type of models, all the necessary equations needed to compute the ML can be found in Casals et al. (2016), Section 5.3.2, where expression (5.50) specifically shows the corresponding log-likelihood function used.

Obviously, the estimation through iterative methods may entail some drawbacks with respect to LS techniques, as computational cost and stability issues. We will give some insight to this in the simulation exercise. In turn, it also has some advantages (which are not particularly analyzed here but are certainly of interest in a real case application) as the possibility of including constraints in the estimation (e.g., $b_2 = b_3$ or $a_{1,1} = 0$ in Equation (28a)) after performing classical statistical inference.

4 Nowcast performance evaluation

In this section we carry out several Monte Carlo simulation exercises in order to test the nowcasting performance of TF-MIDAS against U-MIDAS models for different HF-DGPs. We only compare with U-MIDAS, and not standard MIDAS, because of the following reason. In our procedure, the computational issues grow with k , the sampling frequency between HF and LF. Precisely, for low k , U-MIDAS provides a gain with respect to general MIDAS (e.g., with monthly and quarterly data, see Foroni et al., 2015). This makes U-MIDAS the best competitor of our model.

4.1 Simulation design

The simulation design applied in this paper closely follows that one considered in Foroni et al. (2015), in order to make the results comparable. Moreover, the interested reader can make some indirect comparison with MIDAS as well. We simulate the previous HF-DGPs: (i) VAR(1), (ii) VMA(1) and (ii) VMA(3). We consider a series of parameter combinations that seek to represent a wide range of processes, with different degrees of persistence and correlation between HF and LF variables. Including such a variety of DGPs aims to provide more robustness to the conclusions.

Following Foroni et al. (2015), y_t and x_t are initially simulated for all $t = 1, \dots, (T + ES) \times k$, where t is the HF time index and k denotes the sampling frequency of the LF variable (e.g., for a quarterly observed variable y and a monthly indicator x , $k = 3$).

The number of observations simulated for both variables that are used to estimate the nowcasting models is $T \times k = 300$. In order to perform a nowcast comparison we also consider an Evaluation Sample (ES), so both variables are also simulated $ES \times k$ periods ahead. The size of the ES is set equal to $T/2=50$.

Once all the values for both variables are simulated, the ones corresponding to the LF variable are aggregated. We consider two different aggregation rules, depending if the variable is a stock or a flow. In the former case, we aggregate the values applying the so-called *skip-sampling* procedure, which consists in considering the aggregated values as the values of the LF variable y corresponding only to $t = k, 2k, \dots, (T + ES) \times k$; i.e., $\omega(L)y_t = y_t$ for $t = k, 2k, \dots, (T + ES) \times k$. In the case of a flow variable, we obtain the quarterly aggregated value adding the corresponding monthly values; i.e., $\omega(L)y_t = y_t + y_{t-1} + y_{t-2}$.

In the following lines we present the main characteristics of each DGP considered.

DGP I: HF-VAR(1)

Firstly, we consider a HF bivariate VAR process of order 1, defined by Equation (8) with $\Phi_{1,1} = 1 + \phi_{l,1}L$, $\Phi_{2,2} = 1 + \phi_{h,1}L$, $\Delta_{1,2} = \delta_{l,1}$, $\Pi_{1,2} = 0$ and $\Theta_{1,1} = \Theta_{2,2} = 1$. Error terms $e_{y,t}$ and $e_{x,t}$ are sampled independently from the normal distribution with mean 0 and variance chosen such that the unconditional variance of y is equal to 1, given the specifications of the other parameters.

Parameters $\phi_{l,1}$ and $\phi_{h,1}$, which determine the persistence of both series, are chosen to represent three types of processes with different levels of persistence: low ($\phi_{l,1} = \phi_{h,1} = 0.1$), medium ($\phi_{l,1} = \phi_{h,1} = 0.5$) and high ($\phi_{l,1} = \phi_{h,1} = 0.9$).

Parameter $\delta_{l,1}$ reflects the dependence of variable y on the past value of variable x . Its value is chosen such that stationarity of the series is ensured. Following Foroni et al. (2015), which in turn follows Ghysels and Valkanov (2006), values considered for $\delta_{l,1}$ are 0.1, 0.5 and 1.0. As it is usually assumed in the literature, HF variable affects LF variable, but not the other way round.

DGP II: HF-VMA(1)

The first HF bivariate VMA process is of order 1, inferior to the sampling frequency ($k = 3$), so that in the case of a stock variable, previous quarter LF variable's MA component does not affect its current quarter value. It is defined by Equation (8), with $\Phi_{1,1} = \Phi_{2,2} = 1$, $\Delta_{1,2} = 0$, $\Pi_{1,2} = \pi_{l,1}L$, $\Theta_{1,1} = 1 - \theta_{l,1}L$, and $\Theta_{2,2} = 1 - \theta_{h,1}L$.

We consider different combinations of the parameters values, in order to include a wide variety of DGPs. For θ_l and θ_h we consider values $\{0, -0.5, -0.7, -0.9\}$, for δ_l the values $\{0.1, 0.5, 1\}$ and we set $\delta_h = 0$. We distinguish DGPs with $\theta_l = 0$, i.e., with no MA component, with $\theta_h = 0$, and with both parameters different from 0.

DGP III: HF-VMA(3)

The second HF bivariate VMA process is of order 3, equal to the sampling frequency, so that MA component does have an impact on LF variable's current value. This process is also defined by Equation (8) with $\Phi_{1,1} = \Phi_{2,2} = 1$, $\Delta_{1,2} = 0$, $\Pi_{1,2} = \pi_{l,1}L$ and $\Theta_{2,2} = 1 - \theta_{h,1}L$. However, now $\Theta_{1,1} = 1 - \theta_{l,1}L^3$.

Once again we consider different combinations of the parameters values in order to provide more robustness to the analysis. We consider for θ_l and θ_h the values $\{0, -0.5, -0.7, -0.9\}$, for δ_l the values $\{0.1, 0.5, 1\}$ and θ_h is set to 0.

4.2 Models under comparison

For the reasons mentioned before, we consider U-MIDAS as the reference to evaluate the nowcasting performance of TF-MIDAS model. We then estimate and compare the nowcasting performance of:

- 1) U-MIDAS, defined by the following equation:

$$\omega(L)y_t = \mu_0 + c_1\omega(L)y_{t-k} + \delta(L)x_{t-1} + \epsilon_t, \quad t = k, 2k, \dots, k(T_q - 1); \quad (30)$$

where $\delta(L) = \sum_{j=0}^K \delta_j L^j = \delta_0 + \delta_1 L + \dots + \delta_K L^K$ and in our case $k = 3$.

We do not impose the common factor restriction as in Clements and Galvão (2008). The coefficients $\mu_0, c_1, \delta_0, \dots, \delta_K$ are estimated by LS. We use BIC to determine the lag order K . Following Foroni et al. (2015), we consider two different cases, including (i) $K_{max} = k$, i.e., up to 3 lags of x_t , and (ii) $K_{max} = 4k$, i.e., up to 12 lags of x_t .

Once we have estimated the parameters, U-MIDAS nowcast is computed as,

$$\omega(L) \hat{y}_{T \times k + es \times k \mid T \times k + es \times k - 1} = \hat{\mu}_0 + \hat{c}_1 \omega(L) y_{T \times k + es \times k - k} + \hat{\delta}(L) x_{T \times k + es \times k - 1}, \quad (31)$$

2) TF-MIDAS, defined by the equation:

$$\omega(L) y_t = \beta_0 + \frac{a_1(Z)}{b_1(Z)} x_{t-1} + \frac{a_2(Z)}{b_2(Z)} x_{t-2} + \frac{a_3(Z)}{b_3(Z)} x_{t-3} + \frac{\theta(Z)}{\phi(Z)} a_t, \quad t = k, 2k, \dots, kT_q. \quad (32)$$

Multiple parameter combinations of this model (different combinations of polynomials $a_i(Z)$, $b_i(Z)$, $\phi(Z)$ and $\theta(Z)$ for $i = 0, 1, 2, 3$ for a total of twenty models) are estimated in each step. As in the U-MIDAS case and in order to make the comparison as fair as possible, the model chosen is the one with the least BIC, which is finally used to compute the nowcast. Therefore, the simulation results are not all for cases in which the TF-MIDAS model is correctly specified. In fact, as we will detail in the next section, the correct specification does not occur many times.

The corresponding nowcast is obtained from the following calculation:

$$\begin{aligned} \omega(L) \hat{y}_{kT_q + es \mid kT_q + es - 1} = & \hat{\beta}_0 + \frac{\hat{a}_1(Z)}{\hat{b}_1(Z)} x_{kT_q + es - 1} + \frac{\hat{a}_2(Z)}{\hat{b}_2(Z)} x_{kT_q + es - 2} + \\ & \frac{\hat{a}_3(Z)}{\hat{b}_3(Z)} x_{kT_q + es - 3} + \frac{\hat{\theta}(Z)}{\hat{\phi}(Z)} \hat{a}_{kT_q + es} \quad es = k, 2k, \dots, ES \end{aligned} \quad (33)$$

where $\hat{\theta}(Z) = \hat{\theta}_1 Z + \hat{\theta}_2 Z^2 + \dots + \hat{\theta}_q Z^q$.⁷

4.3 Nowcast performance evaluation

4.3.1 Evaluation procedure

The aggregated simulated values for the LF variable corresponding to the ES are used as the actual values to be compared with the nowcasted values. We assume that the information set available for nowcasting consists in values up to period $(T + es - 1) \times k$, with $es = 1, \dots, ES$, for the LF variable, and up to period $(T + es - 1) \times k + k - 1$ for the HF variable. These values are considered known in each nowcasting exercise and are used to estimate TF-MIDAS and U-MIDAS models.

⁷Notice that the polynomial $\hat{\theta}(Z)$ does not include the unit term as $\hat{a}_{kT_q + es}$ is not known at period $kT_q + es$.

Forecasts of y_t are then computed one LF period ahead (or equivalently k HF periods ahead) for each date in the evaluation sample, conditional on HF information available within the LF forecast period: $\hat{y}_{T \times k + es \times k} \mid T \times k + es \times k - 1$. The corresponding nowcast error is calculated as $\hat{y}_{T \times k + es \times k} \mid T \times k + es \times k - 1 - y_{T \times k + es \times k}$.

In line with the literature, the indicator used to compare the out-of-sample nowcasting performance of the alternative mixed-frequency models is the Mean Square Forecasting Error (MSFE) over the evaluation sample. That is, for each replication r we have:

$$MSFE_r = \frac{1}{ES} \sum_{es=1}^{ES} (\hat{y}_{T \times k + es \times k} \mid T \times k + es \times k - 1 - y_{T \times k + es \times k})^2, \quad (34)$$

where $r = 1, \dots, R$. In this research the total number of replications (R) amounts to 500.

Therefore, as in Foroni et al. (2015), for each replication we need to estimate the model and then obtain the one LF period ahead nowcast 50 times: one for each quarter of the evaluation sample. In each replication the alternative mixed-frequency models are compared through the ratio of their MSFE: $MSFE_{TF-MIDAS} / MSFE_{U-MIDAS}$.

In order to perform the calculations we use the MatLab toolboxes E4 (Casals et al., 2016) and Midas (Ghysels, E. and collaborators, 2017). Main descriptive statistics of this ratio are presented in the following tables for each one of the DGPs considered and the cases of stock and flow variables.

4.3.2 Monte Carlo nowcast comparison results

The relative nowcast performance of TF-MIDAS respect to U-MIDAS is presented in Tables 2-4.

In the case of a HF-VAR(1) DGP, we do not find any advantage in using TF-MIDAS, whether LF variable y is a stock or a flow, considering different parameter combinations and both, $K_{max} = 3$ and $K_{max} = 12$ (see, Tables 2A-2B and Figure 1). This result is expected in the case of a stock variable as TF-MIDAS model expression is observationally equivalent to the corresponding U-MIDAS, including one period-lag of the dependent variable. In the case of a flow variable, one could expect TF-MIDAS would provide some nowcasting performance improvement, as it accommodates an existing MA component to the error term. However, it appears that this effect is so low in this case that no improvement is observed. The MA parameter ψ in TF-MIDAS model is a direct function of the value of the AR parameter ϕ_l in the DGP. Even considering a highly persistent LF process, e.g. $\phi_l = 0.9$, parameter ψ reaches a small value of approximately 0.22. So, most of times, the BIC criterion chooses a TF-MIDAS model that does not include the MA term and the potentially benefit of using this model disappears.⁸

In the case of a HF-VMA(1) DGP, we do not find that TF-MIDAS provides a

⁸However, some alternative simulations have been performed using the AIC criterion to select the best forecasting model. As this criterion overweight the goodness-of-fit against the parsimony relative to BIC, the criterion does choose the TF-MIDAS with an MA term most of the times, leading to the expected nowcasting performance improvement with respect to U-MIDAS models. The results of these simulations are available from the authors upon request.

significant improvement in nowcasting performance relative to U-MIDAS for a stock variable (Table 3A and Figure 1).

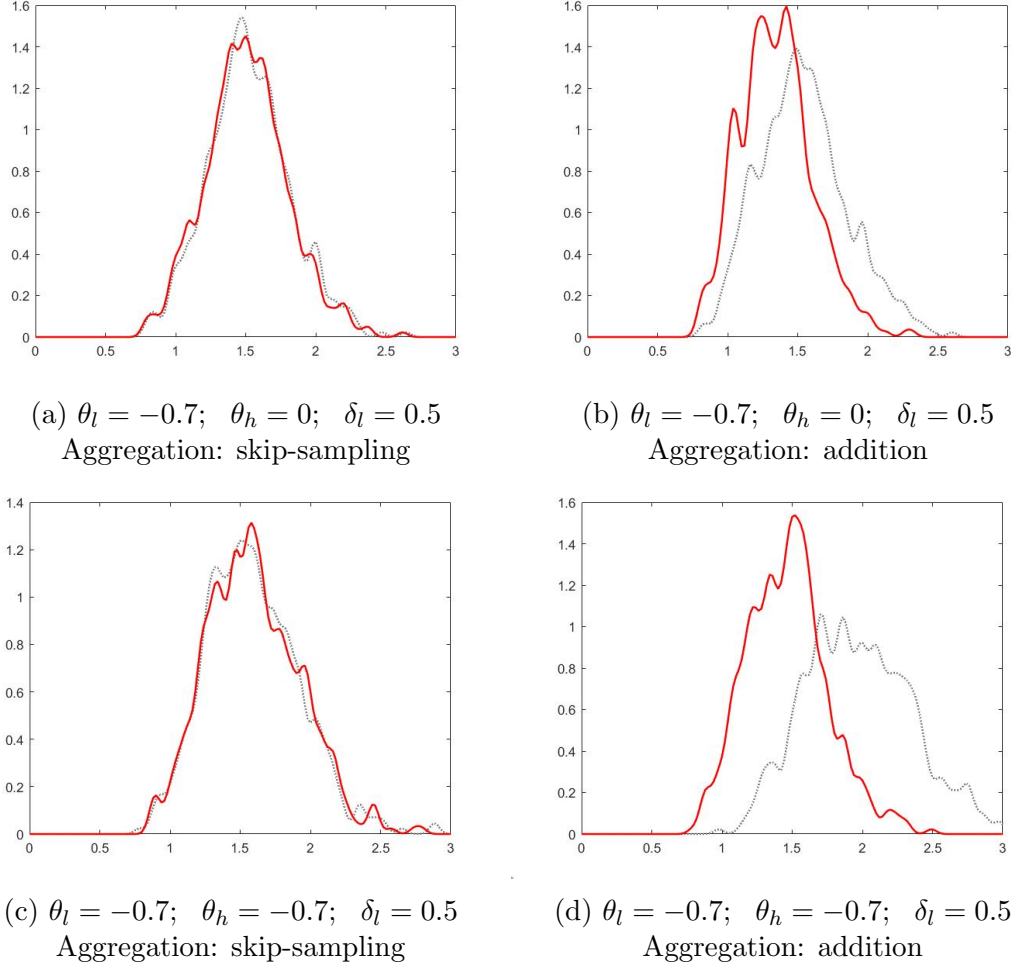


Figure 1: Kernel density estimates for $MSFE_{TF-MIDAS}$ (red) and $MSFE_{U-MIDAS}$ (grey). DGP: HF-VMA(1). $K_{max} = 3$. Rows: same parameter values. Columns: same aggregation scheme.

However, when a flow aggregation scheme is considered, TF-MIDAS shows a significant improvement of its relative nowcasting performance that goes, in median terms, up to 56% when $K_{max} = 3$ and 22% when $K_{max} = 12$ (Table 3B and Figure 1). The explanation of these results mainly comes from the MA structure of the DGP, as we are considering a HF-VMA DGP of order 1, less than the sampling frequency k , which is 3. This causes that in the stock-variable case the previous quarter LF error term ($e_{y,t-3}$) has no effect on the current value of the LF variable (i.e., TF-MIDAS model has no MA component). But in the flow-variable case, when addition is applied as aggregation method and $\theta_l \neq 0$, this effect endures and a non-negligible MA component appears in TF-MIDAS model.

The greater difference in nowcasting performance in favour of TF-MIDAS is observed when TF-MIDAS' two potential sources of relative advantage occur: (i) a polynomial quotient expression, which particularly damages U-MIDAS when K_{max} is short, and, (ii) a

MA component in the mixed-frequency model appears. In the DGP context, this happens for $\theta_l \neq 0$ and $\theta_h \neq 0$. Besides, results show that TF-MIDAS relative performance improves as the value of δ_l and/or θ_l or θ_h , in absolute value, increase (Figure 1b). Lastly, as said before, TF-MIDAS advantage tends to reduce when a greater value K_{max} is chosen (e.g., $K_{max} = 12$). This is a reasonable result, as considering more lags of the indicator makes U-MIDAS model to better approximate TF-MIDAS' coefficient quotients.

Finally, as predicted by theory developed in Section 3.2, in the case of a HF-VMA(3) DGP, TF-MIDAS advantage is observed whether a stock or a flow variable y is considered (Tables 4A-4B and Figure 3). The benefits in terms of relative performance go up to 23% for $K_{max} = 3$ and 30% for $K_{max} = 12$ when the aggregation method is skip-sampling (stock), and to 36% and 19% for $K_{max} = 3$ and $K_{max} = 12$, respectively, when the aggregation method is addition (flow). Notice that for a stock aggregation scheme, the only case where no improvement is reported is when parameter θ_l is zero.

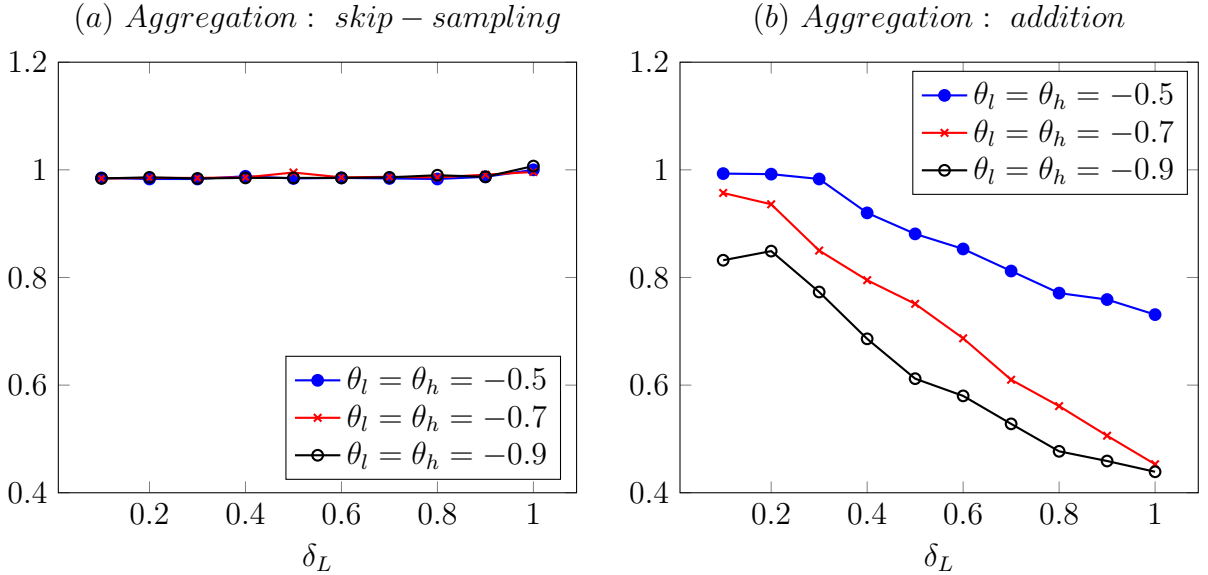
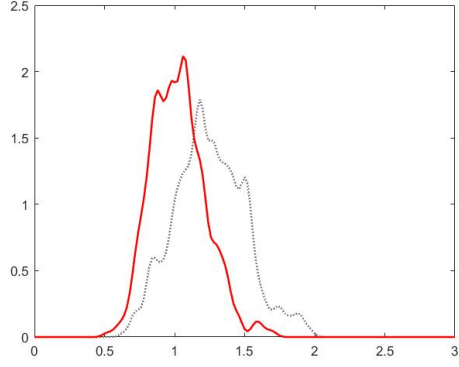
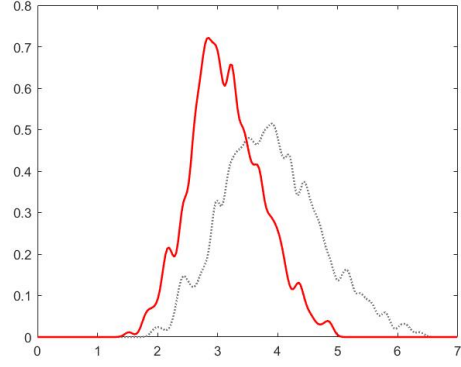


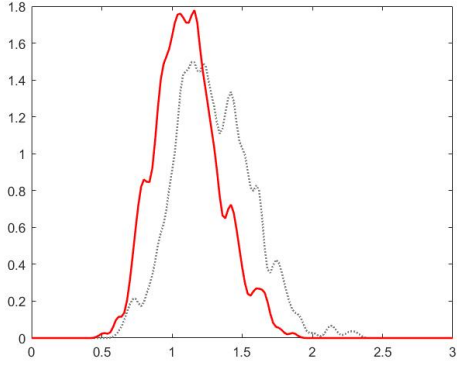
Figure 2: $Median \left(\frac{MSFE_{TF-MIDAS}}{MSFE_{U-MIDAS}} \right)$; DGP: HF-VMA(1); $K_{max} = 3$.



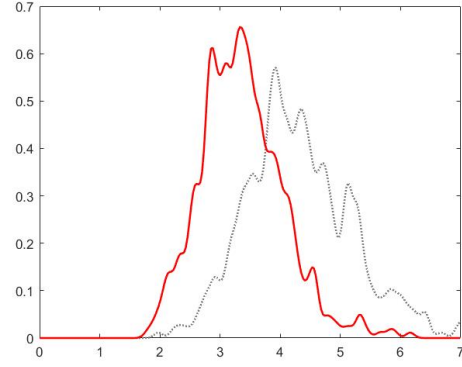
(a) $\theta_l = -0.9$; $\theta_h = 0$; $\delta_l = 0.5$
Aggregation: skip-sampling



(b) $\theta_l = -0.9$; $\theta_h = 0$; $\delta_l = 0.5$
Aggregation: addition



(c) $\theta_l = -0.9$; $\theta_h = -0.9$; $\delta_l = 0.5$
Aggregation: skip-sampling



(d) $\theta_l = -0.9$; $\theta_h = -0.9$; $\delta_l = 0.5$
Aggregation: addition

Figure 3: Kernel density estimates for $MSFE_{TF-MIDAS}$ (red) and $MSFE_{U-MIDAS}$ (grey). DGP: HF-VMA(3). $K_{max} = 3$. Rows: same parameter values. Columns: same aggregation scheme.

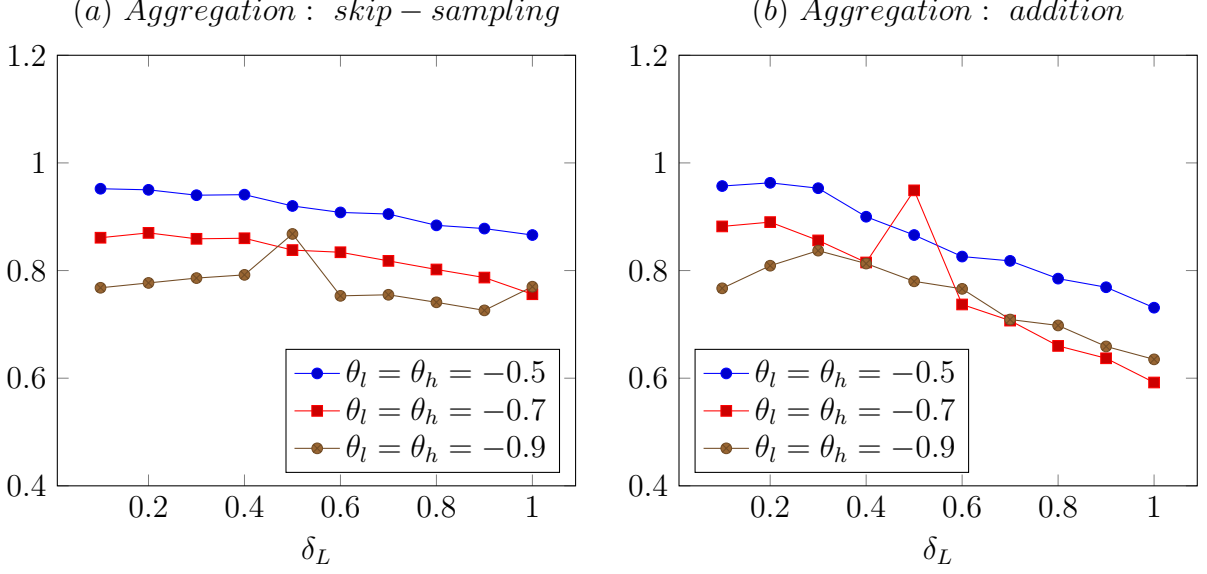


Figure 4: $Median \left(\frac{MSFE_{TF-MIDAS}}{MSFE_{U-MIDAS}} \right)$; DGP: HF-VMA(3); $K_{max} = 3$.

4.3.3 Estimation performance and model specification

This section presents some results on the analysis of the estimation performance and the models selected by BIC in the previous simulations for TF-MIDAS. For the first case, we study the number of times the estimation does not converge. For the second case, we analyze the models chosen by BIC and whether these coincide with the theoretical LF model corresponding to the HF DGP. The later can only be done for TF-MIDAS as the alternatives do not have an exact correspondence between HF and LF models.

We summarize the results in averages, although the variance is important across DGPs (VAR(1), VMA(1) and VMA(3)) and different values for the parameters in the simulation; see Section 4.1.⁹

Most of the times the ML estimation converges. We detect stability problems in VAR(1) and VMA(1) DGPs with the stock aggregation in less than 1% of the simulations. This value increases with the flow aggregation, as this generates more complex LF structures. Particularly, to less than 2% and 5% for VAR(1) and VMA(1) DGPs, respectively. For VMA(3), we find a risk of noninvertibility in models estimated when the parameter $\theta_l = 0.9$. This situation produces an increase in the lack of convergence (specifically affected by some combination of parameter values) that occurs in average in less than 8% of the simulations.

Regarding the specification, not unexpectedly the selection of the models is strongly dependent on the values of the parameters in the DGPs. When the parameters are low (e.g., $\rho = \delta_l = 0.1$, see Section 4.1), BIC shows difficulties to find the correct TF-MIDAS and usually choose less complex models (see discussion in Section 4.3.2).

⁹A detailed analysis is available from the authors upon request.

For low parameter values this happens about 60%-80% of the cases against 5%-25% of correct specifications, the rest being other models. Conversely, when the values of the parameters grow, BIC improves the model selection and the corresponding theoretical TF-MIDAS is chosen in 60-95% of the cases, which benefits the nowcasting results.

5 Conclusions

This paper deals with the mixed-frequency modeling problem from a new perspective. Instead of drawing upon the common distributed lag polynomial model, we employ a transfer function representation to develop a new type of models, named TF-MIDAS.

We describe the model and derive the theoretical TF-MIDAS implied by the HF (bivariate) VARMA family models, depending on two aggregation scheme. This exact correspondence leads to two potential gains in terms of nowcasting and forecasting performance against the U-MIDAS, its best competitor. This is so because TF-MIDAS adds to current U-MIDAS new terms to capture: (i) different infinite LF variable responses to shocks in the HF variable, and (ii) an MA structure in the error term.

The estimation of the suggested model is addressed by reformulating it in its equivalent state space form and is carried out through exact ML. An advantage of these models is that they lie on common transfer functions solid specification and estimation methods (see, e.g., Box and Jenkins, 1976; Casals et al., 2016). In turn, the main downside is that the estimation is computationally more expensive and complex than alternatives based on LS.

As predicted by the theoretical development, an extensive Monte Carlo simulation exercise confirms that TF-MIDAS beats commonly used U-MIDAS models in terms of out-of-sample nowcasting performance for HF-VMA(1) and HF-VMA(3) processes, while their performance remains equal for HF-VAR(1). These results are robust to different parameter specifications.

The theory and simulations conclude that TF-MIDAS is a good competitor in the general MIDAS family. Its characteristics makes it probably the best choice for low values of k (the sampling frequency between HF and LF) and N (the number of HF indicators); e.g., when the indicator is monthly and the output is quarterly sampled. Although this seems like a very particular case, it is the most common situation when nowcasting and forecasting macroeconomic variables, what makes it very relevant for economic climate analysts. In contrast, when k remains low but N grows, although TF-MIDAS performs correctly, U-MIDAS could become a better option due to the computational advantage of LS when more indicators (and parameters) are employed. Finally, for large k and N original MIDAS (with either Exponential Almon or Beta functions) is the best option due to parameter proliferation in U-MIDAS and, specially, in TF-MIDAS.

Table 2: Out-of-sample MSFE of TF-MIDAS relative to U-MIDAS (DGP: HF-VAR(1))

A) Aggregation method: skip-sampling

Parameters			Relative performance ($K_{max}=3$)					Relative performance ($K_{max}=12$)						
ρ	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile
0.1	0.1	0	0.996	0.962	0.982	0.993	1.009	1.030	0.995	0.963	0.982	0.994	1.009	1.029
0.1	0.5	0	0.993	0.958	0.978	0.991	1.008	1.033	0.992	0.955	0.974	0.991	1.009	1.032
0.1	1.0	0	0.996	0.957	0.975	0.995	1.014	1.035	0.995	0.956	0.974	0.993	1.013	1.034
0.5	0.1	0	1.000	0.964	0.981	0.996	1.015	1.041	0.999	0.959	0.980	0.996	1.015	1.039
0.5	0.5	0	1.010	0.952	0.981	1.008	1.038	1.071	1.008	0.949	0.979	1.006	1.036	1.067
0.5	1.0	0	1.023	0.947	0.985	1.020	1.058	1.099	1.019	0.943	0.977	1.017	1.056	1.096
0.9	0.1	0	1.010	0.948	0.976	1.004	1.041	1.085	1.009	0.947	0.976	1.004	1.038	1.084
0.9	0.5	0	1.040	0.961	0.992	1.033	1.075	1.137	1.037	0.958	0.991	1.030	1.074	1.131
0.9	1.0	0	1.285	1.003	1.062	1.187	1.369	1.650	1.281	0.998	1.060	1.179	1.369	1.646

B) Aggregation method: addition

Parameters			Relative performance ($K_{max}=3$)					Relative performance ($K_{max}=12$)								
			ρ	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile
0.1	0.1	0	1.022	0.957	0.986	1.011	1.050	1.100	1.020	0.958	0.984	1.009	1.047	1.096		
0.1	0.5	0	1.043	0.957	0.996	1.033	1.079	1.138	1.038	0.952	0.992	1.031	1.073	1.136		
0.1	1.0	0	1.024	0.947	0.986	1.018	1.051	1.112	1.020	0.940	0.983	1.015	1.050	1.099		
0.5	0.1	0	1.015	0.948	0.977	1.012	1.045	1.090	1.016	0.949	0.981	1.014	1.050	1.090		
0.5	0.5	0	1.027	0.930	0.970	1.018	1.075	1.133	1.022	0.930	0.965	1.014	1.067	1.120		
0.5	1.0	0	0.989	0.872	0.920	0.981	1.046	1.112	1.004	0.916	0.951	0.996	1.054	1.102		
0.9	0.1	0	0.983	0.879	0.925	0.979	1.040	1.099	0.981	0.877	0.922	0.978	1.035	1.095		
0.9	0.5	0	1.004	0.828	0.903	0.992	1.099	1.192	1.041	0.898	0.963	1.031	1.114	1.197		
0.9	1.0	0	1.299	0.745	0.870	1.135	1.598	2.294	1.659	0.970	1.090	1.433	1.963	2.997		

In bold rows where median MSFE of TF-MIDAS is at least 5% better than U-MIDAS.

Table 3: Out-of-sample MSFE of TF-MIDAS relative to U-MIDAS (DGP: HF-VMA(1))

A) Aggregation method: skip-sampling

Parameters				Relative performance ($K_{max}=3$)					Relative performance ($K_{max}=12$)						
θ_l	θ_h	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile
0	-0.5	0.1	0	0.987	0.947	0.968	0.986	1.007	1.030	0.983	0.945	0.965	0.983	1.004	1.024
0	-0.7	0.5	0	1.012	0.962	0.981	1.003	1.037	1.078	0.986	0.936	0.961	0.984	1.009	1.033
0	-0.7	1	0	1.036	0.940	0.980	1.015	1.071	1.154	1.022	0.913	0.964	1.000	1.066	1.147
-0.5	0	0.1	0	0.991	0.956	0.971	0.986	1.010	1.039	0.988	0.949	0.969	0.985	1.007	1.031
-0.7	0	0.5	0	0.990	0.942	0.969	0.988	1.005	1.038	0.986	0.939	0.964	0.986	1.002	1.028
-0.9	0	1	0	0.988	0.945	0.967	0.985	1.005	1.030	0.983	0.943	0.964	0.983	1.002	1.023
-0.5	-0.5	0.1	0	0.988	0.947	0.968	0.985	1.006	1.037	0.982	0.940	0.964	0.981	1.001	1.027
-0.7	-0.7	0.5	0	1.004	0.954	0.974	0.995	1.029	1.071	0.982	0.936	0.960	0.978	1.003	1.032
-0.9	-0.9	1	0	1.016	0.937	0.975	1.007	1.055	1.096	0.993	0.901	0.946	0.985	1.035	1.083

B) Aggregation method: addition

Parameters				Relative performance ($K_{max}=3$)				Relative performance ($K_{max}=12$)							
θ_l	θ_h	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile
0	-0.5	0.1	0	0.996	0.953	0.973	0.989	1.011	1.048	0.988	0.943	0.967	0.986	1.005	1.034
0	-0.7	0.5	0	1.004	0.936	0.972	0.999	1.038	1.103	1.029	0.931	0.971	1.011	1.088	1.160
0	-0.7	1	0	0.798	0.635	0.695	0.781	0.904	1.013	1.010	0.880	0.930	0.983	1.054	1.181
-0.5	0	0.1	0	0.941	0.851	0.886	0.941	0.990	1.031	0.940	0.852	0.891	0.941	0.987	1.024
-0.7	0	0.5	0	0.880	0.779	0.820	0.869	0.934	0.990	0.879	0.780	0.821	0.870	0.931	0.988
-0.9	0	1	0	0.735	0.586	0.631	0.708	0.825	0.936	0.829	0.666	0.718	0.796	0.910	1.050
-0.5	-0.5	0.1	0	0.997	0.940	0.963	0.993	1.020	1.068	0.989	0.933	0.957	0.988	1.012	1.057
-0.7	-0.7	0.5	0	0.752	0.602	0.676	0.751	0.827	0.897	0.910	0.785	0.840	0.904	0.969	1.044
-0.9	-0.9	1	0	0.480	0.291	0.359	0.439	0.586	0.733	0.879	0.573	0.644	0.780	1.093	1.329

In bold rows where median MSFE of TF-MIDAS is at least 5% better than U-MIDAS.

Table 4: Out-of-sample MSFE of TF-MIDAS relative to U-MIDAS (DGP: HF-VMA(3))

A) Aggregation method: skip-sampling

Parameters				Relative performance ($K_{max}=3$)					Relative performance ($K_{max}=12$)						
θ_l	θ_h	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile
0	-0.5	0.1	0	0.998	0.970	0.985	0.994	1.006	1.025	0.997	0.969	0.983	0.993	1.006	1.025
0	-0.7	0.5	0	1.000	0.960	0.980	0.997	1.015	1.041	0.960	0.642	0.760	0.926	1.123	1.324
0	-0.7	1	0	1.010	0.939	0.975	1.004	1.048	1.093	1.069	0.702	0.834	1.020	1.237	1.532
0	-0.9	1	0	0.941	0.814	0.867	0.941	1.005	1.060	0.940	0.819	0.872	0.930	1.011	1.069
-0.5	0	0.1	0	0.954	0.874	0.909	0.950	0.991	1.038	0.953	0.871	0.908	0.950	0.992	1.034
-0.7	0	0.1	0	0.950	0.864	0.905	0.949	0.993	1.037	0.950	0.861	0.904	0.950	0.992	1.038
-0.7	0	0.5	0	0.776	0.656	0.710	0.768	0.830	0.899	0.806	0.696	0.740	0.796	0.863	0.933
-0.9	0	0.5	0	0.828	0.725	0.769	0.816	0.889	0.965	0.827	0.725	0.767	0.818	0.885	0.954
-0.9	0	1	0	0.699	0.577	0.624	0.690	0.763	0.843	0.889	0.552	0.685	0.843	1.043	1.247
-0.5	-0.5	0.1	0	0.956	0.876	0.910	0.952	0.998	1.044	0.955	0.874	0.912	0.952	0.999	1.043
-0.7	-0.7	0.1	0	0.873	0.767	0.812	0.861	0.926	1.000	0.871	0.766	0.808	0.859	0.926	1.000
-0.7	-0.7	0.5	0	0.849	0.731	0.780	0.838	0.912	0.973	0.851	0.735	0.787	0.840	0.908	0.976
-0.9	-0.9	0.5	0	0.856	0.740	0.790	0.850	0.921	0.984	0.860	0.736	0.790	0.852	0.923	0.991
-0.9	-0.9	1	0	0.782	0.658	0.710	0.770	0.851	0.912	0.733	0.488	0.583	0.707	0.856	1.011

B) Aggregation method: addition

Parameters				Relative performance ($K_{max}=3$)					Relative performance ($K_{max}=12$)						
θ_l	θ_h	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile
0	-0.5	0.1	0	0.997	0.946	0.972	0.994	1.020	1.056	0.996	0.942	0.971	0.994	1.020	1.056
0	-0.7	0.5	0	0.941	0.821	0.879	0.944	1.003	1.060	0.972	0.896	0.930	0.973	1.009	1.056
0	-0.7	1	0	0.800	0.650	0.715	0.793	0.878	0.946	1.005	0.910	0.960	1.003	1.052	1.091
0	-0.9	1	0	0.749	0.566	0.647	0.734	0.848	0.954	0.994	0.789	0.854	0.964	1.093	1.250
-0.5	0	0.1	0	1.219	1.039	1.104	1.177	1.298	1.432	1.216	1.033	1.102	1.178	1.293	1.426
-0.7	0	0.1	0	0.953	0.870	0.907	0.946	0.998	1.041	0.952	0.870	0.905	0.946	0.997	1.038
-0.7	0	0.5	0	0.820	0.712	0.754	0.810	0.873	0.955	0.819	0.701	0.744	0.815	0.876	0.954
-0.9	0	0.5	0	0.817	0.697	0.745	0.809	0.879	0.945	0.818	0.702	0.751	0.811	0.874	0.938
-0.9	0	1	0	0.854	0.611	0.682	0.810	0.992	1.166	0.892	0.645	0.724	0.849	1.023	1.198
-0.5	-0.5	0.1	0	0.961	0.861	0.915	0.957	1.004	1.057	0.959	0.860	0.912	0.955	1.003	1.053
-0.7	-0.7	0.1	0	0.888	0.777	0.828	0.882	0.943	1.013	0.887	0.775	0.826	0.882	0.942	1.009
-0.7	-0.7	0.5	0	0.956	0.878	0.908	0.949	0.995	1.037	0.954	0.873	0.908	0.946	0.995	1.037
-0.9	-0.9	0.5	0	0.786	0.648	0.706	0.780	0.860	0.934	0.828	0.691	0.749	0.823	0.901	0.979
-0.9	-0.9	1	0	0.636	0.429	0.519	0.635	0.746	0.851	0.838	0.592	0.668	0.829	0.963	1.101

In bold rows where median MSFE of TF-MIDAS is at least 5% better than U-MIDAS.

References

- Bai, J., Ghysels, E., and Wright, J. H. (2013). State space models and MIDAS regressions. *Econometric Reviews*, 32(7):779–813.
- Box, G. E. P. and Jenkins, G. M. (1976). *Time series analysis: forecasting and control*. Holden-Day.
- Casals, J., García-Hiernaux, A., Jerez, M., and Sotoca, S. (2016). *State-space methods for time series analysis: theory, applications and software*. Chapman & Hall.
- Castle, J. and Hendry, D. (2013). Forecasting and nowcasting macroeconomic variables: A methodological overview. Economics Series Working Paper 674, University of Oxford, Department of Economics.
- Clements, M. P. and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics*, 26(4):546–554.
- Clements, M. P. and Galvão, A. B. (2009). Forecasting US output growth using leading indicators: an appraisal using MIDAS models. *Journal of Applied Econometrics*, 24(7):1187–1206.
- Duarte, C., Rodrigues, P. M. M., and Rua, A. (2017). A mixed frequency approach to the forecasting of private consumption with ATM/POS data. *International Journal of Forecasting*, 33(1):61–75.
- Foroni, C., Marcellino, M., and Schumacher, C. (2012). U-MIDAS: MIDAS regressions with unrestricted lag polynomials. CEPR Discussion Paper 8828.
- Foroni, C., Marcellino, M., and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society Series A*, 178(1):57–82.
- Foroni, C. and Marcellino, M. G. (2013). A survey of econometric methods for mixed-frequency data. Norges Bank Working Paper.
- Garcia-Hiernaux, A., Casals, J., and Jerez, M. (2009). Fast estimation methods for time-series models in state-space form fast estimation methods for time-series models in state-space form. *Journal of Statistical Computation and Simulation*, 79(2):121–134.
- Ghysels, E. (2014). Matlab toolbox for mixed sampling frequency data analysis using MIDAS regression models. Technical report.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2002). The MIDAS touch: Mixed data sampling regression models. Working paper, UNC and UCLA.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2003). Predicting volatility: Getting the most out of return data sampled at different frequencies. Anderson School of Management working paper and UNC Department of Economics working paper.
- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, 76(3):509 – 548.

- Ghysels, E., Santa-Clara, P., and Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1):59–95.
- Ghysels, E., Sinko, A., and Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.
- Ghysels, E. and Valkanov, R. (2006). Linear time series processes with mixed data sampling and MIDAS regression models. *Mimeo*.
- Schumacher, C. (2014). MIDAS regressions with time-varying parameters: An application to corporate bond spreads and GDP in the euro area. Annual conference 2014 (Hamburg): Evidence-based economic policy, Verein für Socialpolitik / German Economic Association.

Appendix

HF-VAR(1) and stock y variable

The equations defining the DGP corresponding to a VAR(1) are the following:

$$y_t = \phi_l y_{t-1} + \delta_l x_{t-1} + e_{y,t} \quad (35a)$$

$$x_t = \phi_h x_{t-1} + e_{x,t} \quad (35b)$$

According to Equation 35a, we can write:

$$y_{t-k} = \phi_l y_{t-k-1} + \delta_l x_{t-k-1} + e_{y,t-k} \quad t = 1, 2, 3, \dots \quad (36)$$

Considering the previous expression for $k = 1$ and $k = 2$, and substituting them in Equation 35a, we obtain:

$$\begin{aligned} y_t &= \phi_l [\phi_l y_{t-2} + \delta_l x_{t-2} + e_{y,t-1}] + \delta_l x_{t-1} + e_{y,t} \\ &= \phi_l^2 y_{t-2} + \phi_l \delta_l x_{t-2} + \phi_l e_{y,t-1} + \delta_l x_{t-1} + e_{y,t} \\ &= \phi_l^2 [\phi_l y_{t-3} + \delta_l x_{t-3} + e_{y,t-2}] + \phi_l \delta_l x_{t-2} + \phi_l e_{y,t-1} + \delta_l x_{t-1} + e_{y,t} \\ &= \phi_l^3 y_{t-3} + \phi_l^2 \delta_l x_{t-3} + \phi_l^2 e_{y,t-2} + \phi_l \delta_l x_{t-2} + \phi_l e_{y,t-1} + \delta_l x_{t-1} + e_{y,t} \end{aligned} \quad (37)$$

Rearranging terms, we obtain the following equation:

$$(1 - \phi_l^3 L^3) y_t = \delta_l x_{t-1} + \phi_l \delta_l x_{t-2} + \phi_l^2 \delta_l x_{t-3} + [1 + \phi_l L + \phi_l^2 L^2] e_{y,t} \quad (38)$$

We can then rewrite the previous equation in TF form, expressing the aggregated value of y (i.e., y_t^A) in terms of HF values of variable x and an error term:

$$\begin{aligned} y_t^A &= \frac{\delta_l}{1 - \phi_l^3 L^3} x_{t-1} + \frac{\phi_l \delta_l}{1 - \phi_l^3 L^3} x_{t-2} + \frac{\phi_l^2 \delta_l}{1 - \phi_l^3 L^3} x_{t-3} + \frac{1 + \phi_l L + \phi_l^2 L^2}{1 - \phi_l^3 L^3} e_{y,t} \\ &= \frac{\delta_l}{1 - \phi_l^3 L^3} x_{t-1} + \frac{\phi_l \delta_l}{1 - \phi_l^3 L^3} x_{t-2} + \frac{\phi_l^2 \delta_l}{1 - \phi_l^3 L^3} x_{t-3} + \frac{1}{1 - \phi_l L} e_{y,t} \quad t = 4, 5, 6, \dots \end{aligned} \quad (39)$$

where to obtain the last expression we have applied that $1 - \phi_l^3 L^3 = (1 - \phi_l L)(1 + \phi_l L + \phi_l^2 L^2)$.

Finally, if we consider the aggregated value of y for each quarter (i.e., $y_{t_q}^A$), from the first expression of the previous equation we can derive that:

$$y_{t_q}^A = \frac{\delta_l}{1 - \phi_l^3 Z} x_{2,t_q} + \frac{\phi_l \delta_l}{1 - \phi_l^3 Z} x_{1,t_q} + \frac{\phi_l^2 \delta_l}{1 - \phi_l^3 Z} x_{3,t_q-1} + \frac{1}{1 - \phi_l^3 Z} \eta_{t_q} \quad t_q = 2, 3, \dots \quad (40)$$

where $y_{t_q}^A$ is the aggregated quarterly value of y , which is a stock variable (i.e. $y_{t_q}^A = y_t$ for $t = 3 * t_q$); x_{2,t_q} is the second monthly value of x for current quarter; x_{1,t_q} is the first monthly value of x for current quarter; x_{3,t_q-1} is the last monthly value of x for the previous quarter; and $\eta_{t_q} = (1 + \phi_l L + \phi_l^2 L^2) e_{y,t}$.

We now calculate the main moments of the error term η_{y^A,t_q} in Equation 40, expressed as a function of $e_{y,t}$.

$$E[\eta_{t_q}] = E[(1 + \phi_l L + \phi_l^2 L^2) e_{y,t}] = 0 \quad (41)$$

$$V[\eta_{t_q}] = V[(1 + \phi_l L + \phi_l^2 L^2) e_{y,t}] = (1 + \phi_l^2 + \phi_l^4) V[e_{y,t}] \quad (42)$$

$$\begin{aligned} Cov(\eta_{t_q}, \eta_{t_q-k}) &= E[\eta_{t_q} * \eta_{t_q-k}] \\ &= E[((1 + \phi_l L + \phi_l^2 L^2) e_{y,t}) * ((1 + \phi_l L + \phi_l^2 L^2) e_{y,t-3k})] \quad (43) \\ &= 0 \quad k = 1, 2, 3, \dots \end{aligned}$$

Therefore we can state that:

$$\rho_k = 0 \quad k = 1, 2, 3, \dots \quad (44)$$

The error term η_{t_q} is then a white noise.

Finally, we can rewrite Equation 40 and express the aggregated value of a stock variable y for each quarter (i.e., $y_{t_q}^A$) in TF form:

$$y_{t_q}^A = \frac{\delta_l}{1 - \phi_l^3 Z} x_{2,t_q} + \frac{\phi_l \delta_l}{1 - \phi_l^3 Z} x_{1,t_q} + \frac{\phi_l^2 \delta_l}{1 - \phi_l^3 Z} x_{3,t_q-1} + \zeta_{t_q} \quad (45a)$$

$$(1 - \phi_l^3 Z) \zeta_{t_q} = e_{y^A,t_q} \quad t_q = 2, 3, \dots \quad (45b)$$

where the error term ζ_{t_q} has an AR(1) structure, and e_{y^A, t_q} is a white noise, uncorrelated with the explicative variables x_{1, t_q} , x_{2, t_q} and x_{3, t_q} , and with this variance:

$$V[e_{y^A, t_q}] = V[\eta_{t_q}] = (1 + \phi_l^2 + \phi_l^4) V[e_{y, t}] \quad (46)$$

HF-VAR(1) and flow y variable

From the expression for y_t obtained in Equation 39 we can deduce equivalent expressions for y_{t-1} and y_{t-2} :

$$y_{t-1} = \frac{\delta_l}{1 - \phi_l^3 L^3} x_{t-2} + \frac{\phi_l \delta_l}{1 - \phi_l^3 L^3} x_{t-3} + \frac{\phi_l^2 \delta_l L^3}{1 - \phi_l^3 L^3} x_{t-1} + \frac{L + \phi_l L^2 + \phi_l^2 L^3}{1 - \phi_l^3 L^3} e_{y, t} \quad (47a)$$

$$y_{t-2} = \frac{\delta_l}{1 - \phi_l^3 L^3} x_{t-3} + \frac{\phi_l \delta_l L^3}{1 - \phi_l^3 L^3} x_{t-1} + \frac{\phi_l^2 \delta_l L^3}{1 - \phi_l^3 L^3} x_{t-2} + \frac{L^2 + \phi_l L^3 + \phi_l^2 L^4}{1 - \phi_l^3 L^3} e_{y, t} \quad (47b)$$

In order to arrive at the previous equations we have recalled that $x_{t-4} = L^3 x_{t-1}$, $x_{t-5} = L^3 x_{t-2}$, $e_{y, t-1} = L e_{y, t}$ and $e_{y, t-2} = L^2 e_{y, t}$.

Adding the previous expressions for y_t , y_{t-1} and y_{t-2} we obtain the aggregated quarterly value of y in the case of a flow variable (i.e., y_t^A):

$$\begin{aligned} y_t^A &= y_t + y_{t-1} + y_{t-2} \\ &= \left[\frac{\delta_l}{1 - \phi_l^3 L^3} x_{t-1} + \frac{\phi_l \delta_l}{1 - \phi_l^3 L^3} x_{t-2} + \frac{\phi_l^2 \delta_l}{1 - \phi_l^3 L^3} x_{t-3} + \frac{1 + \phi_l L + \phi_l^2 L^2}{1 - \phi_l^3 L^3} e_{y, t} \right] \\ &\quad + \left[\frac{\delta_l}{1 - \phi_l^3 L^3} x_{t-2} + \frac{\phi_l \delta_l}{1 - \phi_l^3 L^3} x_{t-3} + \frac{\phi_l^2 \delta_l L^3}{1 - \phi_l^3 L^3} x_{t-1} + \frac{L + \phi_l L^2 + \phi_l^2 L^3}{1 - \phi_l^3 L^3} e_{y, t} \right] \\ &\quad + \left[\frac{\delta_l}{1 - \phi_l^3 L^3} x_{t-3} + \frac{\phi_l \delta_l L^3}{1 - \phi_l^3 L^3} x_{t-1} + \frac{\phi_l^2 \delta_l L^3}{1 - \phi_l^3 L^3} x_{t-2} + \frac{L^2 + \phi_l L^3 + \phi_l^2 L^4}{1 - \phi_l^3 L^3} e_{y, t} \right] \end{aligned} \quad (48)$$

Re-organizing terms we obtain the expression in TF form of the quarterly aggregated value of y (i.e., y_t^A) in the case of a flow variable:

$$\begin{aligned}
y_t^A &= \frac{\delta_l + [\phi_l \delta_l (1 + \phi_l)] L^3}{1 - \phi_l^3 L^3} x_{t-1} + \frac{\delta_l (1 + \phi_l) + (\phi_l^2 \delta_l) L^3}{1 - \phi_l^3 L^3} x_{t-2} \\
&+ \frac{\delta_l (1 + \phi_l + \phi_l^2)}{1 - \phi_l^3 L^3} x_{t-3} + \frac{1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4}{1 - \phi_l^3 L^3} e_{y,t} \\
&= \frac{\delta_l + [\phi_l \delta_l (1 + \phi_l)] L^3}{1 - \phi_l^3 L^3} x_{t-1} + \frac{\delta_l (1 + \phi_l) + (\phi_l^2 \delta_l) L^3}{1 - \phi_l^3 L^3} x_{t-2} \\
&+ \frac{\delta_l (1 + \phi_l + \phi_l^2)}{1 - \phi_l^3 L^3} x_{t-3} + \frac{1 + L + L^2}{1 - \phi_l L} e_{y,t} \quad t = 4, 5, 6, \dots
\end{aligned} \tag{49}$$

Finally, from the first equality in the previous equation we can derive this result that expresses the LF aggregated value of a flow variable y (i.e., $y_{t_q}^A$) in terms of the HF values of x :

$$\begin{aligned}
y_{t_q}^A &= \frac{\delta_l + [\phi_l \delta_l (1 + \phi_l)] Z}{1 - \phi_l^3 Z} x_{2,t_q} + \frac{\delta_l (1 + \phi_l) + (\phi_l^2 \delta_l) Z}{1 - \phi_l^3 Z} x_{1,t_q} \\
&+ \frac{\delta_l (1 + \phi_l + \phi_l^2)}{1 - \phi_l^3 Z} x_{3,t_q-1} + \frac{1}{1 - \phi_l^3 Z} \eta_{t_q} \quad t_q = 2, 3, 4, \dots
\end{aligned} \tag{50}$$

where $y_{t_q}^A$ is the aggregated quarterly value of y , which is a flow variable (i.e. $y_{t_q}^A = y_t + y_{t-1} + y_{t-2}$ for $t = 3 * t_q$); x_{2,t_q} is the second monthly value of x for current quarter; x_{1,t_q} is the first monthly value of x for current quarter; x_{3,t_q-1} is the last monthly value of x for the previous quarter; and $\eta_{t_q} = [1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4] e_{y,t}$.

We now calculate the main moments of the error term η_{t_q} in Equation 50, expressed as a function of $e_{y,t}$:

$$E[(1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4) e_{y,t}] = 0 \tag{51}$$

$$\begin{aligned}
&V[(1 + (1 + \phi_l) L + (1 + \phi_l + \phi_l^2) L^2 + \phi_l (1 + \phi_l) L^3 + \phi_l^2 L^4) e_{y,t}] \\
&= (3 + 4\phi_l + 5\phi_l^2 + 4\phi_l^3 + 3\phi_l^4) V[e_{y,t}]
\end{aligned} \tag{52}$$

$$\begin{aligned}
& Cov[(1 + (1 + \phi_l)L + (1 + \phi_l + \phi_l^2)L^2 + \phi_l(1 + \phi_l)L^3 + \phi_l^2L^4) e_{y,t}, \\
& (1 + (1 + \phi_l)L + (1 + \phi_l + \phi_l^2)L^2 + \phi_l(1 + \phi_l)L^3 + \phi_l^2L^4) e_{y,t-3}] \\
& = E[(1 + (1 + \phi_l)L + (1 + \phi_l + \phi_l^2)L^2 + \phi_l(1 + \phi_l)L^3 + \phi_l^2L^4) e_{y,t} \\
& \quad * (1 + (1 + \phi_l)L + (1 + \phi_l + \phi_l^2)L^2 + \phi_l(1 + \phi_l)L^3 + \phi_l^2L^4) e_{y,t-3}] \\
& = \phi_l(1 + \phi_l) V[e_{y,t-3}] + \phi_l^2(1 + \phi_l) V[e_{y,t-4}] \\
& = \phi_l(1 + \phi_l)^2 V[e_{y,t}]
\end{aligned} \tag{53a}$$

$$\begin{aligned}
& Cov[(1 + (1 + \phi_l)L + (1 + \phi_l + \phi_l^2)L^2 + \phi_l(1 + \phi_l)L^3 + \phi_l^2L^4) e_{y,t}, \\
& (1 + (1 + \phi_l)L + (1 + \phi_l + \phi_l^2)L^2 + \phi_l(1 + \phi_l)L^3 + \phi_l^2L^4) e_{y,t-3k}] = 0
\end{aligned} \tag{53b}$$

$$k = 2, 3, 4, \dots$$

Therefore we can state that:

$$\rho_1 = \frac{\phi_l(1 + \phi_l)^2}{3 + 4\phi_l + 5\phi_l^2 + 4\phi_l^3 + 3\phi_l^4} \tag{54a}$$

$$\rho_k = 0 \quad k = 2, 3, 4, \dots \tag{54b}$$

Then, as η_{t_q} has a MA(1) structure, we can state that:

$$\eta_{t_q} = (1 + \psi Z) e_{y^A, t_q} \tag{55}$$

where e_{y^A, t_q} is a white noise.

We can derive the value of ψ in Equation 55, recalling that the acf for a MA(1) model has the following expression:

$$\rho_1 = \frac{\psi}{1 + \psi^2} \tag{56a}$$

$$\rho_k = 0 \quad k = 2, 3, \dots \tag{56b}$$

From Equation 56a, we can derive the value of ψ :

$$\rho_1\psi^2 - \psi + \rho_1 = 0 \tag{57}$$

Solving the previous equation we arrive at this result:

$$\psi = \frac{1 \pm \sqrt{1 - 4\rho_1^2}}{2\rho_1} \quad (58)$$

Between the two solutions for Equation 58, we consider the one which satisfies the invertibility condition: $|\psi| < 1$.

Therefore, we can finally state that the LF aggregated value of a flow variable y (i.e., $y_{t_q}^A$) can be expressed as a model in TF form in the following way:

$$\begin{aligned} y_{t_q}^A &= \frac{\delta_l + [\phi_l \delta_l (1 + \phi_l)]Z}{1 - \phi_l^3 Z} x_{2,t_q} + \frac{\delta_l (1 + \phi_l) + (\phi_l^2 \delta_l)Z}{1 - \phi_l^3 Z} x_{1,t_q} \\ &+ \frac{\delta_l (1 + \phi_l + \phi_l^2)}{1 - \phi_l^3 Z} x_{3,t_q-1} + \zeta_{t_q} \quad t_q = 2, 3, 4, \dots \end{aligned} \quad (59a)$$

$$(1 - \phi_l^3 Z) \zeta_{t_q} = (1 + \psi Z) e_{y^A, t_q} \quad (59b)$$

where the error term ζ_{t_q} has an ARMA(1,1) structure, and e_{y^A, t_q} is a white noise, uncorrelated with the explicative variables x_{1,t_q} , x_{2,t_q} and x_{3,t_q} , with the following variance.

From Equations 52 and 55 we can derive an expression for $V[e_{y^A, t_q}]$:

$$\begin{aligned} V[\eta_{t_q}] &= V[(1 + \psi Z) e_{y^A, t_q}] = V[e_{y^A, t_q} + \psi e_{y^A, t_q-1}] \\ &= V[e_{y^A, t_q}] + \psi^2 V[e_{y^A, t_q-1}] = (1 + \psi^2) V[e_{y^A, t_q}] \end{aligned} \quad (60)$$

Therefore we can conclude that:

$$V[e_{y^A, t_q}] = \frac{V[\eta_{t_q}]}{1 + \psi^2} = \frac{(3 + 4\phi_l + 5\phi_l^2 + 4\phi_l^3 + 3\phi_l^4)}{1 + \psi^2} V[e_{y, t}] \quad (61)$$

HF-VMA(1) and stock y variable

The equations defining the DGP corresponding to a VMA(1), where we assume again for simplicity that $\delta_h = 0$ [i.e., past values of y do not affect the present value of x], are the following:

$$y_t = e_{y, t} + \theta_l e_{y, t-1} + \delta_l e_{x, t-1} \quad (62a)$$

$$x_t = e_{x, t} + \theta_h e_{x, t-1} \quad (62b)$$

From Equation 62b we can state that:

$$x_{t-k} = e_{x, t-k} + \theta_h e_{x, t-k-1} \quad k = 1, 2, 3, \dots \quad (63)$$

And from this equation we can derive that:

$$e_{x,t-k} = x_{t-k} - \theta_h e_{x,t-k-1} \quad k = 1, 2, \dots \quad (64)$$

Taking into account Equation 64, we deduce that $e_{x,t-1} = x_{t-1} / (1 + \theta_h L)$. Re-arranging terms, we can rewrite Equation 62a as follows:

$$\begin{aligned} y_t &= e_{y,t} + \theta_l e_{y,t-1} + \delta_l \left[\frac{x_{t-1}}{1 + \theta_h L} \right] \\ &= \frac{\delta_l}{1 + \theta_h L} x_{t-1} + (1 + \theta_l L) e_{y,t} \end{aligned} \quad (65)$$

where we have applied that $(1 + \theta_h L)(1 + \theta_l L) = 1 + (\theta_h + \theta_l) L + (\theta_h \theta_l) L^2$.

If we develop the first term of the previous equation according a geometric progression with common ratio $1 + \theta_h L$, we can rewrite Equation 65 in the following way:

$$\begin{aligned} y_t^A &= \delta_l \sum_{i=0}^{\infty} (-\theta_h^3 L^3)^i x_{t-1} - \delta_l \theta_h \sum_{i=0}^{\infty} (-\theta_h^3 L^3)^i x_{t-2} \\ &\quad + \delta_l \theta_h^2 \sum_{i=0}^{\infty} (-\theta_h^3 L^3)^i x_{t-3} + (1 + \theta_l L) e_{y,t} \\ &= \frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-1} - \frac{\delta_l \theta_h}{1 + \theta_h^3 L^3} x_{t-2} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 L^3} x_{t-3} + (1 + \theta_l L) e_{y,t} \\ &\quad t = 4, 5, \dots \end{aligned} \quad (66)$$

Considering the aggregated value of the stock variable y for each quarter (i.e., $y_{t_q}^A$), we can rewrite the previous equation in this way:

$$y_{t_q}^A = \frac{\delta_l}{1 + \theta_h^3 Z} x_{2,t_q} - \frac{\delta_l \theta_h}{1 + \theta_h^3 Z} x_{1,t_q} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 Z} x_{3,t_q-1} + \eta_{t_q} \quad t_q = 2, 3, \dots \quad (67)$$

where $y_{t_q}^A$ is the aggregated quarterly value of y , which is a stock variable (i.e. $y_{t_q}^A = y_t$ for $t = 3 * t_q$); x_{2,t_q} is the second monthly value of x for current quarter; x_{1,t_q} is the first monthly value of x for current quarter; x_{3,t_q-1} is the last monthly value of x for the previous quarter; and $\eta_{t_q} = (1 + \theta_l Z) e_{y,t}$.

We now calculate the main moments of the error term η_{t_q} , expressed as a function of $e_{y,t}$:

$$E[\eta_{t_q}] = E[(1 + \theta_l L) e_{y,t}] = 0 \quad (68)$$

$$V[\eta_{t_q}] = V[(1 + \theta_l L) e_{y,t}] = V[e_{y,t}] + \theta_l^2 V[e_{y,t-1}] = (1 + \theta_l^2) V[e_{y,t}] \quad (69)$$

$$\begin{aligned} Cov(\eta_{t_q}, \eta_{t_q-k}) &= E[\eta_{t_q} * \eta_{t_q-k}] = E[(1 + \theta_l L) e_{y,t} * (1 + \theta_l L) e_{y,t-3k}] \\ &= 0 \quad k = 1, 2, \dots \end{aligned} \quad (70)$$

Thus we can state:

$$\rho_k = 0 \quad k = 1, 2, 3, \dots \quad (71)$$

Finally, we can conclude that the expression for the aggregated quarterly value of y in the case of a stock variable is:

$$y_{t_q}^A = \frac{\delta_l}{1 + \theta_h^3 Z} x_{2,t_q} - \frac{\delta_l \theta_h}{1 + \theta_h^3 Z} x_{1,t_q} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 Z} x_{3,t_q-1} + \zeta_{t_q} \quad (72a)$$

$$\zeta_{t_q} = e_{y^A,t_q} \quad t_q = 2, 3, 4, \dots \quad (72b)$$

where the error term ζ_{t_q} is a white noise, and e_{y^A,t_q} is a white noise, uncorrelated with the explicative variables x_{1,t_q} , x_{2,t_q} and x_{3,t_q} , with variance: $V[e_{y^A,t_q}] = V[\eta_{t_q}] = (1 + \theta_l^2) V[e_{y,t}]$.

HF-VMA(1) and flow y variable

From the expression for y_t obtained in Equation 66 we can deduce equivalent expressions for y_{t-1} and y_{t-2} :

$$y_{t-1} = \frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-2} - \frac{\delta_l \theta_h}{1 + \theta_h^3 L^3} x_{t-3} + \frac{\delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-1} + (L + \theta_l L^2) e_{y,t} \quad (73a)$$

$$y_{t-2} = \frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-3} - \frac{\delta_l \theta_h L^3}{1 + \theta_h^3 L^3} x_{t-1} + \frac{\delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-2} + (L^2 + \theta_l L^3) e_{y,t} \quad (73b)$$

In order to arrive at the previous equations we have recalled that $x_{t-4} = L^3 x_{t-1}$, $x_{t-5} = L^3 x_{t-2}$, $e_{y,t-1} = L e_{y,t}$ and $e_{y,t-2} = L^2 e_{y,t}$.

Adding the previous expressions for y_t , y_{t-1} and y_{t-2} we obtain the aggregated

value of y (i.e., y_t^A) in the case of a flow variable:

$$\begin{aligned}
y_t^A &= y_t + y_{t-1} + y_{t-2} \\
&= \left[\frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-1} - \frac{\delta_l \theta_h}{1 + \theta_h^3 L^3} x_{t-2} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 L^3} x_{t-3} + (1 + \theta_l L) e_{y,t} \right] \\
&\quad + \left[\frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-2} - \frac{\delta_l \theta_h}{1 + \theta_h^3 L^3} x_{t-3} + \frac{\delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-1} + (L + \theta_l L^2) e_{y,t} \right] \\
&\quad + \left[\frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-3} - \frac{\delta_l \theta_h L^3}{1 + \theta_h^3 L^3} x_{t-1} + \frac{\delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-2} + (L^2 + \theta_l L^3) e_{y,t} \right]
\end{aligned} \tag{74}$$

Re-organizing terms we obtain the expression in TF form of the quarterly aggregated value of a flow variable y :

$$\begin{aligned}
y_t^A &= \frac{\delta_l + \delta_l \theta_h (\theta_h - 1) L^3}{1 + \theta_h^3 L^3} x_{t-1} + \frac{\delta_l (1 - \theta_h) + \delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-2} + \frac{\delta_l (1 - \theta_h + \theta_h^2)}{1 + \theta_h^3 L^3} x_{t-3} \\
&\quad + (1 + \theta_l L)(1 + L + L^2) e_{y,t} \quad t = 4, 5, \dots
\end{aligned} \tag{75}$$

where we have applied that $1 + (1 + \theta_l)L + (1 + \theta_l)L^2 + \theta_l L^3 = (1 + \theta_l L)(1 + L + L^2)$.

If we consider the aggregated value of the flow variable y for each quarter (i.e., $y_{t_q}^A$), we can rewrite the previous equation in this way:

$$\begin{aligned}
y_{t_q}^A &= \frac{\delta_l + \delta_l \theta_h (\theta_h - 1) Z}{1 + \theta_h^3 Z} x_{2,t_q} + \frac{\delta_l (1 - \theta_h) + \delta_l \theta_h^2 Z}{1 + \theta_h^3 Z} x_{1,t_q} \\
&\quad + \frac{\delta_l (1 - \theta_h + \theta_h^2)}{1 + \theta_h^3 Z} x_{3,t_q-1} + \eta_{t_q} \quad t_q = 2, 3, \dots
\end{aligned} \tag{76}$$

where $y_{t_q}^A$ is the aggregated quarterly value of y , which is a stock variable (i.e. $y_{t_q}^A = y_t$ for $t = 3 * t_q$); x_{2,t_q} is the second monthly value of x for current quarter; x_{1,t_q} is the first monthly value of x for current quarter; x_{3,t_q-1} is the last monthly value of x for the previous quarter; and $\eta_{t_q} = [1 + (1 + \theta_l)L + (1 + \theta_l)L^2 + \theta_l L^3] e_{y,t}$.

We now calculate the main moments of the error term η_{t_q} , expressed as a function of $e_{y,t}$.

$$E[\eta_{t_q}] = E[(1 + (1 + \theta_l)L + (1 + \theta_l)L^2 + \theta_l L^3) e_{y,t}] = 0 \tag{77}$$

$$\begin{aligned}
V[\eta_{t_q}] &= V[(1 + (1 + \theta_l)L + (1 + \theta_l)L^2 + \theta_l L^3) e_{y,t}] \\
&= V[e_{y,t}] + (1 + \theta_l)^2 V[e_{y,t-1}] + (1 + \theta_l)^2 V[e_{y,t-2}] + \theta_l^2 V[e_{y,t-3}] \quad (78) \\
&= [3 + 4\theta_l + 3\theta_l^2] V[e_{y,t}]
\end{aligned}$$

$$\begin{aligned}
Cov(\eta_{t_q}, \eta_{t_{q-1}}) &= E[\eta_{t_q} * \eta_{t_{q-1}}] \\
&= E[(1 + (1 + \theta_l)L + (1 + \theta_l)L^2 + \theta_l L^3) e_{y,t} * (1 + (1 + \theta_l)L + (1 + \theta_l)L^2 + \theta_l L^3) e_{y,t-3}] \\
&= E[(e_{y,t} + (1 + \theta_l) e_{y,t-1} + (1 + \theta_l) e_{y,t-2} + \theta_l e_{y,t-3}) \\
&\quad * (e_{y,t-3} + (1 + \theta_l) e_{y,t-4} + (1 + \theta_l) e_{y,t-5} + \theta_l e_{y,t-6})] \\
&= \theta_l V[e_{y,t-3}] \quad (79a)
\end{aligned}$$

$$\begin{aligned}
Cov(\eta_{t_q}, \eta_{t_{q-k}}) &= E[\eta_{t_q} * \eta_{t_{q-k}}] \\
&= E[(e_{y,t} + (1 + \theta_l) e_{y,t-1} + (1 + \theta_l) e_{y,t-2} + \theta_l e_{y,t-3}) \\
&\quad * (e_{y,t-3k} + (1 + \theta_l) e_{y,t-3k-1} + (1 + \theta_l) e_{y,t-3k-2} + \theta_l e_{y,t-3k-3})] \\
&= 0 \quad k = 2, 3, \dots \quad (79b)
\end{aligned}$$

From the previous equations we can state that:

$$\rho_1 = \frac{\theta_l V[e_{y,t}]}{(3 + 4\theta_l + 3\theta_l^2) V[e_{y,t}]} = \frac{\theta_l}{3 + 4\theta_l + 3\theta_l^2} \quad (80a)$$

$$\rho_k = 0 \quad k = 1, 2, 3, \dots \quad (80b)$$

Therefore we can deduce that the error term η_{t_q} has a MA(1) structure.

Finally, we can conclude that the expression for the aggregated quarterly value of y in the case of a flow variable is:

$$y_{t_q}^A = \frac{\delta_l + \delta_l \theta_h (\theta_h - 1) Z}{1 + \theta_h^3 Z} x_{2,t_q} + \frac{\delta_l (1 - \theta_h) + \delta_l \theta_h^2 Z}{1 + \theta_h^3 Z} x_{1,t_q} + \frac{\delta_l (1 - \theta_h + \theta_h^2)}{1 + \theta_h^3 Z} x_{3,t_q-1} + \zeta_{t_q} \quad (81a)$$

$$\zeta_{t_q} = (1 + \psi Z) e_{y^A,t_q} \quad t_q = 2, 3, 4, \dots \quad (81b)$$

where the error term ζ_{t_q} has a MA(1) structure, and e_{y^A,t_q} is a white noise, uncorrelated with the explicative variables x_{1,t_q} , x_{2,t_q} and x_{3,t_q} . Coefficient ψ satisfies this condition:

$$\psi = \frac{1 \pm \sqrt{1 - 4\rho_1^2}}{2\rho_1} \quad (82)$$

Between the two solutions for Equation 82, we consider the one which satisfies the invertibility condition: $|\psi| < 1$.

The variance of e_{y^A,t_q} can be derived from Equations 78 and 81b, considering that $\eta_{t_q} = \zeta_{t_q}$:

$$\begin{aligned} V[\eta_{t_q}] &= V[(1 + \psi Z) e_{y^A,t_q}] = V[e_{y^A,t_q} + \psi e_{y^A,t_q-1}] \\ &= V[e_{y^A,t_q}] + \psi^2 V[e_{y^A,t_q-1}] = (1 + \psi^2) V[e_{y^A,t_q}] \end{aligned} \quad (83)$$

Therefore we can conclude that:

$$V[e_{y^A,t_q}] = \frac{V[\eta_{t_q}]}{1 + \psi^2} = \frac{3 + 4\theta_l + 3\theta_l^2}{1 + \psi^2} V[e_{y,t}] \quad (84)$$

HF-VMA(3) and stock y variable

The equations defining the DGP corresponding to a VMA(3) are the following:

$$y_t = e_{y,t} + \theta_l e_{y,t-3} + \delta_l e_{x,t-1} \quad (85a)$$

$$x_t = e_{x,t} + \theta_h e_{x,t-1} \quad (85b)$$

From Equation 85a we can state that:

$$y_{t-3} = e_{y,t-3} + \theta_l e_{y,t-6} + \delta_l e_{x,t-4} \quad (86)$$

And from the previous equation we can derive this result:

$$e_{y,t-3} = y_{t-3} - \theta_l e_{y,t-6} - \delta_l e_{x,t-4} \quad (87)$$

Substituting the previous expression in Equation 85a we obtain:

$$y_t = e_{y,t} + \theta_l y_{t-3} - \theta_l^2 e_{y,t-6} + \delta_l(1 - \theta_l L^3) e_{x,t-1} \quad (88)$$

From Equation 85b we can derive that:

$$e_{x,t-1} = x_{t-1} - \theta_h e_{x,t-2} \quad k = 1, 2, 3, \dots \quad (89)$$

Substituting the previous result in Equation 88 we obtain this:

$$\begin{aligned} y_t &= e_{y,t} + \theta_l y_{t-3} - \theta_l^2 e_{y,t-6} + \delta_l(1 - \theta_l L^3) e_{x,t-1} \\ &= \theta_l y_{t-3} + (1 - \theta_l^2 L^6) e_{y,t} + \delta_l(1 - \theta_l L^3) \left[\frac{x_{t-1}}{1 + \theta_h L} \right] \end{aligned} \quad (90)$$

We can rewrite Equation 90 in the following equivalent way:

$$\begin{aligned} y_t &= \theta_l y_{t-3} + \frac{\delta_l(1 - \theta_l L^3)}{1 + \theta_h^3 L^3} x_{t-1} - \frac{\delta_l \theta_h(1 - \theta_l L^3)}{1 + \theta_h^3 L^3} x_{t-2} + \frac{\delta_l \theta_h^2(1 - \theta_l L^3)}{1 + \theta_h^3 L^3} x_{t-3} \\ &\quad + \frac{1 + \theta_h L - \theta_l^2 L^6 - \theta_h \theta_l^2 L^7}{1 + \theta_h L} e_{y,t} \quad t = 4, 5, \dots \end{aligned} \quad (91)$$

We can then express the equation in TF form:

$$\begin{aligned} y_t &= \frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-1} - \frac{\delta_l \theta_h}{1 + \theta_h^3 L^3} x_{t-2} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 L^3} x_{t-3} + (1 + \theta_l L^3) e_{y,t} \\ &\quad t = 4, 5, \dots \end{aligned} \quad (92)$$

where we have applied that: $(1 - \theta_l L^3)(1 + \theta_h L)(1 + \theta_l L^3) = 1 + \theta_h L - \theta_l^2 L^6 - \theta_h \theta_l^2 L^7$.

Considering the aggregated value of a stock variable y for each quarter (i.e., $y_{t_q}^A$), we can re-write the previous equation in this way:

$$y_{t_q}^A = \frac{\delta_l}{1 + \theta_h^3 Z} x_{2,t_q} - \frac{\delta_l \theta_h}{1 + \theta_h^3 Z} x_{1,t_q} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 Z} x_{3,t_q-1} + (1 + \theta_l Z) e_{y,t_q} \quad t_q = 2, 3, \dots \quad (93)$$

where $y_{t_q}^A$ is the aggregated quarterly value of y , which is a stock variable (i.e. $y_{t_q}^A = y_t$ for $t = 3 * t_q$); x_{2,t_q} is the second monthly value of x for current quarter; x_{1,t_q} is the

first monthly value of x for current quarter; x_{3,t_q-1} is the last monthly value of x for the previous quarter; and e_{y^A,t_q} is a white noise.

Therefore, we can finally state that the LF aggregated value of a stock variable y (i.e., $y_{t_q}^A$) can be expressed as a model in TF form in the following way:

$$y_{t_q}^A = \frac{\delta_l}{1 + \theta_h^3 Z} x_{2,t_q} - \frac{\delta_l \theta_h}{1 + \theta_h^3 Z} x_{1,t_q} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 Z} x_{3,t_q-1} + \zeta_{t_q} \quad t_q = 2, 3, \dots \quad (94a)$$

$$\zeta_{t_q} = (1 + \theta_l Z) e_{y^A,t_q} \quad (94b)$$

where the error term ζ_{t_q} has a MA(1) structure, and e_{y^A,t_q} is a white noise, uncorrelated with the explicative variables x_{1,t_q} , x_{2,t_q} and x_{3,t_q} , and with variance $V[e_{y^A,t_q}] = V[e_{y,t}]$.

HF-VMA(3) and flow y variable

From the expression for y_t obtained in Equation 92 we can derive equivalent expressions for y_{t-1} and y_{t-2} :

$$y_{t-1} = \frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-2} - \frac{\delta_l \theta_h}{1 + \theta_h^3 L^3} x_{t-3} + \frac{\delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-1} + (L + \theta_l L^4) e_{y,t} \quad (95a)$$

$$y_{t-2} = \frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-3} - \frac{\delta_l \theta_h L^3}{1 + \theta_h^3 L^3} x_{t-1} + \frac{\delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-2} + (L^2 + \theta_l L^5) e_{y,t} \quad (95b)$$

In order to arrive at the previous equations we have recalled that $x_{t-4} = L^3 x_{t-1}$, $x_{t-5} = L^3 x_{t-2}$, $e_{y,t-1} = L e_{y,t}$ and $e_{y,t-2} = L^2 e_{y,t}$.

Adding the previous expressions for y_t , y_{t-1} and y_{t-2} we obtain the aggregated value of y (i.e., y_t^A):

$$\begin{aligned}
y_t^A &= y_t + y_{t-1} + y_{t-2} \\
&= \left[\frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-1} - \frac{\delta_l \theta_h}{1 + \theta_h^3 L^3} x_{t-2} + \frac{\delta_l \theta_h^2}{1 + \theta_h^3 L^3} x_{t-3} + (1 + \theta_l L^3) e_{y,t} \right] \\
&+ \left[\frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-2} - \frac{\delta_l \theta_h}{1 + \theta_h^3 L^3} x_{t-3} + \frac{\delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-1} + (L + \theta_l L^4) e_{y,t} \right] \\
&+ \left[\frac{\delta_l}{1 + \theta_h^3 L^3} x_{t-3} - \frac{\delta_l \theta_h L^3}{1 + \theta_h^3 L^3} x_{t-1} + \frac{\delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-2} + (L^2 + \theta_l L^5) e_{y,t} \right]
\end{aligned} \tag{96}$$

Re-organizing terms we obtain the expression in TF form of the quarterly aggregated value of y in the case of a flow variable:

$$\begin{aligned}
y_t^A &= \frac{\delta_l + \delta_l \theta_h (\theta_h - 1) L^3}{1 + \theta_h^3 L^3} x_{t-1} + \frac{\delta_l (1 - \theta_h) + \delta_l \theta_h^2 L^3}{1 + \theta_h^3 L^3} x_{t-2} + \frac{\delta_l (1 - \theta_h + \theta_h^2)}{1 + \theta_h^3 L^3} x_{t-3} \\
&+ (1 + \theta_l L^3)(1 + L + L^2) e_{y,t} \quad t = 4, 5, \dots
\end{aligned} \tag{97}$$

Finally, if we consider the aggregated value of y for each quarter (i.e., $y_{t_q}^A$), we can re-write the previous equation in this way:

$$\begin{aligned}
y_{t_q}^A &= \frac{\delta_l + \delta_l \theta_h (\theta_h - 1) Z}{1 + \theta_h^3 Z} x_{t-1} - \frac{\delta_l (1 - \theta_h) + \delta_l \theta_h^2 Z}{1 + \theta_h^3 Z} x_{t-2} \\
&+ \frac{\delta_l (1 - \theta_h + \theta_h^2)}{1 + \theta_h^3 Z} x_{t-3} + (1 + \theta_l Z) \eta_{t_q} \quad t = 2, 3, \dots
\end{aligned} \tag{98}$$

where $y_{t_q}^A$ is the aggregated quarterly value of y , which is a flow variable (i.e., $y_{t_q}^A = y_t + y_{t-1} + y_{t-2}$ for $t = 3 * t_q$); x_{2,t_q} is the second monthly value of x for current quarter; x_{1,t_q} is the first monthly value of x for current quarter; x_{3,t_q-1} is the last monthly value of x for the previous quarter; and $\eta_{t_q} = (1 + L + L^2) e_{y,t}$.

We calculate in the following lines the main moments of the error term η_{t_q} , expressed as a function of $e_{y,t}$.

$$E[\eta_{t_q}] = E[(1 + L + L^2) e_{y,t}] = 0 \tag{99}$$

$$V[\eta_{t_q}] = V[(1 + L + L^2) e_{y,t}] = 3 V[e_{y,t}] \tag{100}$$

$$\begin{aligned}
Cov(\eta_{t_q}, \eta_{t_q-k}) &= E[\eta_{t_q} * \eta_{t_q-k}] = E[(1 + L + L^2) e_{y,t} * (1 + L + L^2) e_{y,t-3k}] \\
&= E[(e_{y,t} + e_{y,t-1} + e_{y,t-2}) * (e_{y,t-3k} + e_{y,t-3k-1} + e_{y,t-3k-2})] = 0 \\
& \qquad \qquad \qquad k = 1, 2, 3, \dots
\end{aligned} \tag{101}$$

Then we can derive that:

$$\rho_k = 0 \qquad k = 1, 2, 3, \dots \tag{102}$$

Therefore we can conclude that the expression for the aggregated quarterly value of y in the case of a flow variable is:

$$y_{t_q}^A = \frac{\delta_l + \delta_l \theta_h (\theta_h - 1) Z}{1 + \theta_h^3 Z} x_{2,t_q} + \frac{\delta_l (1 - \theta_h) + \delta_l \theta_h^2 Z}{1 + \theta_h^3 Z} x_{3,t_q} + \frac{\delta_l (1 - \theta_h + \theta_h^2)}{1 + \theta_h^3 Z} x_{1,t_q-1} + \zeta_{t_q} \tag{103a}$$

$$\zeta_{t_q} = (1 + \theta_l Z) e_{y^A, t_q} \qquad t_q = 2, 3, 4, \dots \tag{103b}$$

where the error term ζ_{t_q} has a MA(1) structure, and e_{y^A, t_q} is a white noise, uncorrelated with the explicative variables x_{1,t_q} , x_{2,t_q} and x_{3,t_q} , with variance:

$$V[e_{y^A, t_q}] = V[\eta_{t_q}] = 3 V[e_{y,t}] \tag{104}$$

Additional Tables

Table 5: Out-of-sample MSFE of TF-MIDAS relative to U-MIDAS (DGP: HF-VMA(1); aggregation method: skip-sampling)

Parameters				Relative performance ($K_{max}=3$)					Relative performance ($K_{max}=12$)						
θ_l	θ_h	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile
0	-0.5	0.1	0	0.987	0.947	0.968	0.986	1.007	1.030	0.983	0.945	0.965	0.983	1.004	1.024
0	-0.5	0.5	0	0.986	0.944	0.969	0.986	1.004	1.028	0.985	0.939	0.967	0.985	1.004	1.031
0	-0.5	1	0	1.041	0.943	0.977	1.005	1.078	1.188	1.038	0.934	0.973	1.004	1.080	1.190
0	-0.7	0.1	0	0.999	.948	0.967	0.985	1.008	1.037	0.998	0.946	0.967	0.984	1.007	1.037
0	-0.7	0.5	0	1.012	0.962	0.981	1.003	1.037	1.078	0.986	0.936	0.961	0.984	1.009	1.033
0	-0.7	1	0	1.036	0.940	0.980	1.015	1.071	1.154	1.022	0.913	0.964	1.000	1.066	1.147
0	-0.9	0.1	0	0.990	0.955	0.969	0.987	1.006	1.033	0.989	0.952	0.969	0.986	1.005	1.034
0	-0.9	0.5	0	0.987	0.938	0.963	0.985	1.009	1.038	0.985	0.925	0.958	0.982	1.008	1.042
0	-0.9	1	0	0.983	0.849	0.927	0.987	1.047	1.100	0.997	0.851	0.917	0.986	1.062	1.147
-0.5	0	0.1	0	0.991	0.955	0.971	0.986	1.010	1.039	0.988	0.949	0.969	0.985	1.007	1.031
-0.5	0	0.5	0	0.983	0.940	0.966	0.984	1.000	1.023	0.982	0.938	0.966	0.983	0.999	1.022
-0.5	0	1	0	0.982	0.935	0.965	0.984	1.001	1.028	0.982	0.935	0.963	0.984	1.000	1.028
-0.7	0	0.1	0	0.981	0.940	0.964	0.984	0.998	1.018	0.981	0.940	0.963	0.983	0.999	1.018
-0.7	0	0.5	0	0.990	0.942	0.969	0.988	1.005	1.038	0.986	0.939	0.964	0.986	1.002	1.028
-0.7	0	1	0	0.984	0.944	0.965	0.983	1.001	1.023	0.982	0.941	0.964	0.983	1.002	1.023
-0.9	0	0.1	0	0.983	0.942	0.965	0.982	0.998	1.026	0.982	0.941	0.963	0.982	0.998	1.027
-0.9	0	0.5	0	0.981	0.940	0.963	0.983	0.998	1.024	0.980	0.935	0.961	0.981	0.996	1.023
-0.9	0	1	0	0.988	0.945	0.967	0.985	1.005	1.030	0.983	0.943	0.964	0.983	1.002	1.023
-0.5	-0.5	0.1	0	0.988	0.947	0.968	0.985	1.006	1.037	0.982	0.940	0.964	0.981	1.001	1.027
-0.5	-0.5	0.5	0	0.985	0.948	0.967	0.984	1.001	1.025	0.982	0.942	0.961	0.983	1.002	1.024
-0.5	-0.5	1	0	1.027	0.939	0.972	1.000	1.070	1.150	1.025	0.932	0.966	1.001	1.071	1.151
-0.7	-0.7	0.1	0	0.985	0.946	0.969	0.984	1.002	1.030	0.984	0.942	0.967	0.983	1.000	1.027
-0.7	-0.7	0.5	0	1.004	0.954	0.974	0.995	1.029	1.071	0.982	0.936	0.960	0.978	1.003	1.032
-0.7	-0.7	1	0	1.018	0.939	0.963	0.996	1.045	1.116	1.018	0.923	0.958	0.995	1.055	1.139
-0.9	-0.9	0.1	0	0.984	0.946	0.966	0.984	1.001	1.026	0.983	0.942	0.963	0.982	1.001	1.027
-0.9	-0.9	0.5	0	0.985	0.940	0.963	0.985	1.007	1.034	0.981	0.928	0.959	0.982	1.004	1.035
-0.9	-0.9	1	0	1.016	0.937	0.975	1.007	1.055	1.096	0.993	0.901	0.946	0.985	1.035	1.083

In bold rows where median MSFE of TF-MIDAS is at least 5% better than U-MIDAS.

Table 6: Out-of-sample MSFE of TF-MIDAS relative to U-MIDAS (DGP: HF-VMA(1); aggregation method: addition)

Parameters				Relative performance ($K_{max}=3$)						Relative performance ($K_{max}=12$)					
θ_l	θ_h	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile
0	-0.5	0.1	0	0.996	0.953	0.973	0.989	1.011	1.048	0.988	0.943	0.967	0.986	1.005	1.034
0	-0.5	0.5	0	0.970	0.876	0.923	0.971	1.017	1.068	0.978	0.909	0.941	0.976	1.011	1.055
0	-0.5	1	0	0.886	0.746	0.817	0.883	0.958	1.019	0.983	0.921	0.953	0.981	1.011	1.047
0	-0.7	0.1	0	0.992	0.941	0.968	0.993	1.021	1.044	0.992	0.937	0.966	0.993	1.020	1.048
0	-0.7	0.5	0	1.004	0.936	0.972	0.999	1.038	1.103	1.029	0.931	0.971	1.011	1.088	1.160
0	-0.7	1	0	0.798	0.635	0.695	0.781	0.904	1.013	1.010	0.880	0.930	0.983	1.054	1.181
0	-0.9	0.1	0	0.997	0.948	0.970	0.994	1.018	1.051	0.996	0.943	0.970	0.993	1.018	1.054
0	-0.9	0.5	0	0.932	0.794	0.860	0.938	1.000	1.067	0.962	0.859	0.910	0.962	1.011	1.065
0	-0.9	1	0	0.750	0.547	0.630	0.739	0.852	0.961	1.008	0.785	0.869	0.964	1.097	1.295
-0.5	0	0.1	0	0.941	0.851	0.886	0.941	0.990	1.031	0.940	0.852	0.891	0.941	0.987	1.024
-0.5	0	0.5	0	0.957	0.876	0.918	0.954	0.991	1.038	0.953	0.870	0.914	0.953	0.988	1.033
-0.5	0	1	0	0.937	0.846	0.891	0.938	0.978	1.035	0.932	0.852	0.889	0.930	0.970	1.019
-0.7	0	0.1	0	0.943	0.849	0.889	0.937	0.994	1.045	0.941	0.848	0.885	0.935	0.992	1.044
-0.7	0	0.5	0	0.880	0.779	0.820	0.869	0.934	0.990	0.879	0.780	0.821	0.870	0.931	0.988
-0.7	0	1	0	0.830	0.718	0.763	0.822	0.886	0.955	0.884	0.766	0.812	0.877	0.939	1.004
-0.9	0	0.1	0	0.806	0.693	0.741	0.805	0.875	0.990	0.805	0.692	0.740	0.805	0.871	0.982
-0.9	0	0.5	0	0.812	0.633	0.695	0.779	0.915	1.032	0.830	0.653	0.712	0.801	0.932	1.052
-0.9	0	1	0	0.735	0.586	0.631	0.708	0.825	0.936	0.829	0.666	0.718	0.796	0.910	1.050
-0.5	-0.5	0.1	0	0.997	0.940	0.963	0.993	1.020	1.068	0.989	0.933	0.957	0.988	1.012	1.057
-0.5	-0.5	0.5	0	0.882	0.761	0.812	0.881	0.937	1.005	0.935	0.853	0.894	0.935	0.973	1.015
-0.5	-0.5	1	0	0.731	0.580	0.647	0.731	0.803	0.875	0.958	0.854	0.901	0.950	1.009	1.067
-0.7	-0.7	0.1	0	0.956	0.863	0.906	0.957	0.998	1.051	0.955	0.863	0.906	0.956	0.998	1.049
-0.7	-0.7	0.5	0	0.752	0.602	0.676	0.751	0.827	0.897	0.910	0.785	0.840	0.904	0.969	1.044
-0.7	-0.7	1	0	0.466	0.361	0.401	0.453	0.521	0.598	0.864	0.739	0.796	0.855	0.923	0.989
-0.9	-0.9	0.1	0	0.838	0.723	0.772	0.832	0.889	0.990	0.837	0.724	0.771	0.831	0.889	0.988
-0.9	-0.9	0.5	0	0.635	0.462	0.526	0.612	0.733	0.835	0.771	0.566	0.642	0.753	0.885	0.997
-0.9	-0.9	1	0	0.480	0.291	0.359	0.439	0.586	0.733	0.879	0.573	0.644	0.780	1.093	1.329

In bold rows where median MSFE of TF-MIDAS is at least 5% better than U-MIDAS.

Table 7: Out-of-sample MSFE of TF-MIDAS relative to U-MIDAS (DGP: HF-VMA(3); aggregation method: skip-sampling)

Parameters				Relative performance ($K_{max}=3$)						Relative performance ($K_{max}=12$)					
θ_l	θ_h	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile
0	-0.5	0.1	0	0.998	0.970	0.985	0.994	1.006	1.025	0.997	0.969	0.983	0.993	1.006	1.025
0	-0.5	0.5	0	0.986	0.945	0.967	0.986	1.005	1.031	0.984	0.937	0.961	0.984	1.003	1.033
0	-0.5	1	0	1.034	0.942	0.971	0.998	1.064	1.187	1.031	0.936	0.965	0.995	1.060	1.189
0	-0.7	0.1	0	0.987	0.944	0.965	0.985	1.004	1.031	0.986	0.945	0.963	0.985	1.004	1.033
0	-0.7	0.5	0	1.000	0.960	0.980	0.997	1.015	1.041	0.960	0.642	0.760	0.926	1.123	1.324
0	-0.7	1	0	1.010	0.939	0.975	1.004	1.048	1.093	1.069	0.702	0.834	1.020	1.237	1.532
0	-0.9	0.1	0	0.985	0.944	0.967	0.985	1.005	1.041	0.984	0.943	0.966	0.984	1.007	1.039
0	-0.9	0.5	0	0.989	0.939	0.963	0.986	1.009	1.044	0.988	0.927	0.958	0.984	1.011	1.051
0	-0.9	1	0	0.941	0.814	0.867	0.941	1.005	1.060	0.940	0.819	0.872	0.930	1.011	1.069
-0.5	0	0.1	0	0.954	0.874	0.909	0.950	0.991	1.038	0.953	0.871	0.908	0.950	0.992	1.034
-0.5	0	0.5	0	0.924	0.828	0.872	0.919	0.975	1.022	0.922	0.827	0.869	0.914	0.974	1.022
-0.5	0	1	0	0.985	0.838	0.897	0.971	1.055	1.147	1.073	0.876	0.952	1.052	1.168	1.305
-0.7	0	0.1	0	0.950	0.864	0.905	0.949	0.993	1.037	0.950	0.861	0.904	0.950	0.992	1.038
-0.7	0	0.5	0	0.776	0.656	0.710	0.768	0.830	0.899	0.806	0.696	0.740	0.796	0.863	0.933
-0.7	0	1	0	0.774	0.667	0.706	0.770	0.838	0.922	0.808	0.694	0.739	0.806	0.872	0.960
-0.9	0	0.1	0	2.135	0.642	0.684	0.739	0.797	0.861	2.126	0.640	0.685	0.740	0.796	0.859
-0.9	0	0.5	0	0.828	0.725	0.769	0.816	0.889	0.965	0.827	0.725	0.767	0.818	0.885	0.954
-0.9	0	1	0	0.699	0.577	0.624	0.690	0.763	0.843	0.889	0.552	0.685	0.843	1.043	1.247
-0.5	-0.5	0.1	0	0.956	0.876	0.910	0.952	0.998	1.044	0.955	0.874	0.912	0.952	0.999	1.043
-0.5	-0.5	0.5	0	0.927	0.829	0.877	0.920	0.973	1.028	0.926	0.838	0.876	0.913	0.974	1.029
-0.5	-0.5	1	0	0.874	0.770	0.818	0.866	0.922	0.980	0.912	0.799	0.844	0.907	0.970	1.025
-0.7	-0.7	0.1	0	0.873	0.767	0.812	0.861	0.926	1.000	0.871	0.766	0.808	0.859	0.926	1.000
-0.7	-0.7	0.5	0	0.849	0.731	0.780	0.838	0.912	0.973	0.851	0.735	0.787	0.840	0.908	0.976
-0.7	-0.7	1	0	0.762	0.639	0.693	0.756	0.829	0.898	0.848	0.707	0.773	0.841	0.920	1.000
-0.9	-0.9	0.1	0	0.807	0.647	0.689	0.747	0.811	0.876	0.806	0.647	0.687	0.746	0.812	0.875
-0.9	-0.9	0.5	0	0.856	0.740	0.790	0.850	0.921	0.984	0.860	0.736	0.790	0.852	0.923	0.991
-0.9	-0.9	1	0	0.782	0.658	0.710	0.770	0.851	0.912	0.733	0.488	0.583	0.707	0.856	1.011

In bold rows where median MSFE of TF-MIDAS is at least 5% better than U-MIDAS.

Table 8: Out-of-sample MSFE of TF-MIDAS relative to U-MIDAS (DGP: HF-VMA(3); aggregation method: addition)

Parameters				Relative performance ($K_{max}=3$)						Relative performance ($K_{max}=12$)					
θ_l	θ_h	δ_l	δ_h	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile	Mean	10th percentile	25th percentile	Median	75th percentile	90th percentile
0	-0.5	0.1	0	0.997	0.946	0.972	0.994	1.020	1.056	0.996	0.942	0.971	0.994	1.020	1.056
0	-0.5	0.5	0	1.010	0.866	0.920	0.967	1.013	1.079	1.025	0.902	0.936	0.973	1.005	1.056
0	-0.5	1	0	0.884	0.749	0.813	0.879	0.948	1.027	0.988	0.920	0.953	0.987	1.017	1.062
0	-0.7	0.1	0	0.997	0.941	0.967	0.995	1.024	1.058	0.995	0.933	0.967	0.995	1.022	1.061
0	-0.7	0.5	0	0.941	0.821	0.879	0.944	1.003	1.060	0.972	0.896	0.930	0.973	1.009	1.056
0	-0.7	1	0	0.800	0.650	0.715	0.793	0.878	0.946	1.005	0.910	0.960	1.003	1.052	1.091
0	-0.9	0.1	0	0.992	0.941	0.965	0.991	1.017	1.048	0.991	0.940	0.966	0.990	1.017	1.049
0	-0.9	0.5	0	0.923	0.789	0.857	0.936	0.999	1.055	0.966	0.865	0.916	0.970	1.020	1.070
0	-0.9	1	0	0.749	0.566	0.647	0.734	0.848	0.954	0.994	0.789	0.854	0.964	1.093	1.250
-0.5	0	0.1	0	1.219	1.039	1.104	1.177	1.298	1.432	1.216	1.033	1.102	1.178	1.293	1.426
-0.5	0	0.5	0	0.919	0.811	0.860	0.915	0.969	1.025	0.913	0.805	0.856	0.910	0.968	1.020
-0.5	0	1	0	0.935	0.854	0.898	0.934	0.975	1.015	1.269	0.992	1.096	1.227	1.418	1.565
-0.7	0	0.1	0	0.953	0.870	0.907	0.946	0.998	1.041	0.952	0.870	0.905	0.946	0.997	1.038
-0.7	0	0.5	0	0.820	0.712	0.754	0.810	0.873	0.955	0.819	0.701	0.744	0.815	0.876	0.954
-0.7	0	1	0	0.786	0.661	0.714	0.770	0.845	0.926	0.814	0.682	0.734	0.798	0.876	0.962
-0.9	0	0.1	0	0.758	0.655	0.700	0.755	0.825	0.955	0.757	0.655	0.700	0.755	0.822	0.948
-0.9	0	0.5	0	0.817	0.697	0.745	0.809	0.879	0.945	0.818	0.702	0.751	0.811	0.874	0.938
-0.9	0	1	0	0.854	0.611	0.682	0.810	0.992	1.166	0.892	0.645	0.724	0.849	1.023	1.198
-0.5	-0.5	0.1	0	0.961	0.861	0.915	0.957	1.004	1.057	0.959	0.860	0.912	0.955	1.003	1.053
-0.5	-0.5	0.5	0	0.865	0.743	0.803	0.866	0.925	0.978	0.891	0.799	0.838	0.885	0.943	0.997
-0.5	-0.5	1	0	0.741	0.603	0.655	0.731	0.812	0.897	0.900	0.795	0.843	0.898	0.954	1.007
-0.7	-0.7	0.1	0	0.888	0.777	0.828	0.882	0.943	1.013	0.887	0.775	0.826	0.882	0.942	1.009
-0.7	-0.7	0.5	0	0.956	0.878	0.908	0.949	0.995	1.037	0.954	0.873	0.908	0.946	0.995	1.037
-0.7	-0.7	1	0	0.609	0.467	0.528	0.592	0.668	0.766	0.863	0.705	0.764	0.840	0.919	1.005
-0.9	-0.9	0.1	0	1.778	0.681	0.710	0.767	0.835	0.899	1.781	0.680	0.709	0.765	0.834	0.902
-0.9	-0.9	0.5	0	0.786	0.648	0.706	0.780	0.860	0.934	0.828	0.691	0.749	0.823	0.901	0.979
-0.9	-0.9	1	0	0.636	0.429	0.519	0.635	0.746	0.851	0.838	0.592	0.668	0.829	0.963	1.101

In bold rows where median MSFE of TF-MIDAS is at least 5% better than U-MIDAS.